HANDBOOK OF COMBINATORICS

Volume 2

edited by

R.L. GRAHAM
AT&T Bell Laboratories, Murray Hill, NJ, USA

M. GRÖTSCHEL
Konrad-Zuse-Zentrum für Informationstechnik, Berlin-Wilmersdorf, Germany

L. LOVÁSZ
Yale University, New Haven, CT, USA

1995

ELSEVIER
AMSTERDAM - LAUSANNE - NEW YORK
OXFORD - SHANNON - TOKYO

ELSEVIER SCIENCE B.V.
Sara Burgerhartstraat 25
P.O. Box 211, 1000 AE Amsterdam, The Netherlands

Co-publishers for the United States and Canada:

The MIT Press
55 Hayward Street
Cambridge, MA 02142, U.S.A.

Elsevier Science B.V.
ISBN: 0-444-82351-4 (Volume 2)
ISBN: 0-444-88002-X (Set of vols 1 and 2)

The MIT Press
ISBN: 0-262-07171-1 (Volume 2)
ISBN: 0-262-07169-X (Set of vols 1 and 2)

No responsibility is assumed by the publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein.

This book is printed on acid-free paper.

Printed in The Netherlands

Preface

Combinatorics belongs to those areas of mathematics having experienced a most impressive growth in recent years. This growth has been fuelled in large part by the increasing importance of computers, the needs of computer science and demands from applications where discrete models play more and more important roles. But also more classical branches of mathematics have come to recognize that combinatorial structures are essential components of many mathematical theories.

Despite the dynamic state of this development, we feel that the time is ripe for summarizing the current status of the field and for surveying those major results that in our opinion will be of long-term importance. We approached leading experts in all areas of combinatorics to write chapters for this Handbook. The response was overwhelmingly enthusiastic and the result is what you see here.

The intention of the Handbook is to provide the working mathematician or computer scientist with a good overview of basic methods and paradigms, as well as important results and current issues and trends across the broad spectrum of combinatorics. However, our hope is that even specialists can benefit from reading this Handbook, by learning a leading expert's coherent and individual view of the topic.

As the reader will notice by looking at the table of contents, we have structured the Handbook into five sections: Structures, Aspects, Methods, Applications, and Horizons. We feel that viewing the whole field from different perspectives and taking different cross-sections will help to understand the underlying framework of the subject and to see the interrelationships more clearly. As a consequence of this approach, a number of the fundamental results occur in more than one chapter. We believe that this is an asset rather than a shortcoming, since it illustrates different viewpoints and interpretations of the results.

We thank the authors not only for writing the chapters but also for many helpful suggestions on the organization of the book and the presentation of the material. Many colleagues have contributed to the Handbook by reading the initial versions of the chapters and by making proposals with respect to the inclusion of topics and results as well as the structuring of the chapters. We are grateful for the significant help we received.

Even though this Handbook is quite voluminous, it was inevitable that some areas of combinatorics had to be left out or were not covered in the depth they deserved. Nevertheless, we believe that the Handbook of Combinatorics presents a comprehensive and accessible view of the present state of the field and that it will prove to be of lasting value.

Ronald Graham
Martin Grötschel
László Lovász

Contents

Volume I

Graphs

Finite Sets and Relations

Matroids

Symmetric Structures

Combinatorial Structures in Geometry and Number Theory

Volume II

Part II: Aspects 1019

Part III: Methods 1647

Part IV: Applications 1873

Part V: Horizons 2083

List of Contributors

Alon, N., *Tel Aviv University, Tel Aviv* (Ch. 32).

Babai, L., *Eötvös University, Budapest and University of Chicago, Chicago, IL* (Ch. 27).

Beck, J., *Rutgers University, New Brunswick, NJ* (Ch. 26).

Biggs, N.L., *University of London, London* (Ch. 44).

Bixby, R.E., *Rice University, Houston, TX* (Ch. 11).

Björner, A., *Royal Institute of Technology, Stockholm* (Ch. 34).

Bollobás, B., *University of Cambridge, Cambridge and Louisiana State University, Baton Rouge, LA* (Ch. 23).

Bondy, J.A., *Université Claude Bernard Lyon 1, Villeurbanne and University of Waterloo, Waterloo, Ont.* (Ch. 1).

Brouwer, A.E., *Eindhoven University of Technology, Eindhoven* (Chs. 14, 15).

Cameron, P.J., *Queen Mary and Westfield College, London* (Chs. 12, 13).

Cunningham, W.H., *University of Waterloo, Waterloo, Ont.* (Ch. 11).

Duchet, P., *IMAG, Laboratoire de Structures Discrètes et de Didactique, Grenoble Cedex* (Ch. 7).

Erdős, P., *Hungarian Academy of Sciences, Budapest* (Ch. 17).

Frank, A., *Eőtvős University, Budapest* (Ch. 2).

Frankl, P., *CNRS, Paris* (Ch. 24).

Gessel, I.M., *Brandeis University, Waltham, MA* (Ch. 21).

Godsil, C.D., *University of Waterloo, Waterloo, Ont.* (Chs. 31, 37).

Grötschel, M., *Konrad-Zuse-Zentrum für Informationstechnik, Berlin-Wilmersdorf* (Chs. 28, 37).

Guy, R.K., *University of Calgary, Calgary, Alta.* (Ch. 43).

Haemers, W.H., *Tilburg University, Tilburg* (Ch. 15).

Hajnal, A., *Hungarian Academy of Sciences, Budapest* (Ch. 42).

Karoński, M., *Adam Mickiewicz University, Poznań and Emory University, Atlanta, GA* (Ch. 6).

Klee, V., *University of Washington, Seattle, WA* (Ch. 18).

Kleinschmidt, P., *Universität Passau, Passau* (Ch. 18).

Kolen, A.W.J., *University of Limburg, Maastricht* (Ch. 35).

Lagarias, J.C., *AT&T Bell Laboratories, Murray Hill, NJ* (Ch. 19).

Lenstra, J.K., *Eindhoven University of Technology, Eindhoven and Centre for Mathematics and Computer Science, Amsterdam* (Ch. 35).

Lloyd, E.K., *University of Southampton, Southampton* (Ch. 44).

Lovász, L., *Yale University, New Haven, CT* (Chs. 28, 31, 40, 41).

Nešetřil, J., *Charles University, Prague* (Ch. 25).

Odlyzko, A.M., *AT&T Bell Laboratories, Murray Hill, NJ* (Ch. 22).

Pomerance, C., *University of Georgia, Athens, GA* (Ch. 20).

Pulleyblank, W.R., *IBM, Thomas J. Watson Research Center, Yorktown Heights, NY* (Ch. 3).

Purdy, G., *University of Cincinnati, Cincinnati, OH* (Ch. 17).

Pyber, L., *Hungarian Academy of Sciences, Budapest* (Ch. 41).

Recski, A., *Technical University of Budapest, Budapest* (Ch. 36).

Rouvray, D.H., *The University of Georgia, Athens, GA* (Ch. 38).

Sárközy, A., *Hungarian Academy of Sciences, Budapest* (Ch. 20).

Schrijver, A., *Centrum voor Wiskunde en Informatica, Amsterdam* (Ch. 30).

Seymour, P.D., *Bell Communications Research, Morristown, NJ* (Appendix to Ch. 4, Ch. 10).

Shmoys, D.B., *Cornell University, Ithaca, NY* (Chs. 29, 40).

Sós, V.T., *Hungarian Academy of Sciences, Budapest* (Ch. 26).

Spencer, J., *New York University, New York, NY* (Ch. 33).

Stanley, R.P., *Massachusetts Institute of Technology, Cambridge, MA* (Ch. 21).

Tardos, É., *Cornell University, Ithaca, NY* (Chs. 29, 40).

Thomassen, C., *The Technical University of Denmark, Lyngby* (Ch. 5).

Toft, B., *Odense University, Odense* (Ch. 4).

Trotter, W.T., *Arizona State University, Tempe, AZ* (Ch. 8).

Van Lint, J.H., *Eindhoven University of Technology, Eindhoven* (Ch. 16).

Waterman, M.S., *University of Southern California, Los Angeles, CA* (Ch. 39).

Welsh, D.J.A., *University of Oxford, Oxford* (Chs. 9, 37, 41).

Wilson, R.J., *The Open University, Milton Keynes* (Ch. 44).

Ziegler, G.M., *Konrad-Zuse-Zentrum für Informationstechnik, Berlin-Wilmersdorf* (Ch. 41).

Part II
Aspects

CHAPTER 21

Algebraic Enumeration

Ira M. GESSEL*

Department of Mathematics, Brandeis University, Waltham, MA 02254, USA

Richard P. STANLEY**

Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Contents

*Partially supported by NSF Grant DMS-8902666.
**Partially supported by NSF Grant DMS-8401376.

HANDBOOK OF COMBINATORICS
Edited by R. Graham, M. Grötschel and L. Lovász

List of symbols and definitions

Symbol or definition	*Meaning*	*Page*
$[n]$	$\{1,2,\ldots,n\}$	1023
composition		1023
functional digraph		1024
left–right minimum		1025
lower record		1025
$c(n,k)$	unsigned Stirling number of the first kind	1027
free monoid		1027
A^*	free monoid generated by A	1027
conjugate of a word		1030
$[x^n]F(x)$	coefficient of x^n in $F(x)$	1032
ordered tree		1033
ballot number		1033
Catalan number		1033
binary tree		1034
Runyon number		1034
Narayana number		1034
transfer matrix method		1035
multiset		1036
partition		1037
$(a)_n$	$(1-a)(1-aq)\cdots(1-aq^{n-1})$	1038
$\left[{n \atop k}\right]$	q-binomial coefficient	1039
exponential generating function		1040
$S(n,k)$	Stirling number of the second kind	1041
B_n	Bell number	1041
descent (of a permutation)		1042
$A_n(t)$	Eulerian polynomial	1042
up–down permutation		1043
D_n	tangent and secant numbers	1043
derangement		1044
descent set		1050
$\mathscr{I}(P)$	incidence algebra of P	1052
ζ	zeta function	1052
μ	Möbius function	1052
lattice		1054
$\vee$	join	1054
meet		1054
$\hat{0}$	minimal element of a lattice	1054
$\hat{1}$	maximal element of a lattice	1054
atom		1054
geometric lattice		1055
symmetric function		1056
Λ^k	symmetric functions of degree k	1056
m_λ	monomial symmetric function	1056
e_λ	elementary symmetric function	1056
h_λ	complete symmetric function	1056
p_λ	power sum symmetric function	1056
s_λ	Schur function	1057
Ferrers diagram		1057
plane partition		1057
hook length		1058
$Z(G)$	cycle index	1059

1. Introduction

In all areas of mathematics, questions arise of the form "Given a description of a finite set, what is its cardinality?" Enumerative combinatorics deals with questions of this sort in which the sets to be counted have a fairly simple structure, and come in indexed families, where the index set is most often the set of nonnegative integers. The two branches of enumerative combinatorics discussed in this book are asymptotic enumeration and algebraic enumeration. In asymptotic enumeration, the basic goal is an approximate but simple formula which describes the order of growth of the cardinalities as a function of their parameters. Algebraic enumeration deals with exact results, either explicit formulas for the numbers in question, or more often, generating functions or recurrences from which the numbers can be computed.

The two fundamental tools in enumeration are *bijections* and *generating functions*, which we introduce in the next two sections. If there is a simple formula for the cardinality of a set, we would like to find a "reason" for the existence of such a formula. For example, if a set S has cardinality 2^n, we may hope to prove this by finding a bijection between S and the set of subsets of an n-element set. The method of generating functions has a long history, but has often been regarded as an ad hoc device. One of the main themes of this article is to explain how generating functions arise naturally in enumeration problems.

Further information and references on the topics discussed here may be found in the books of Comtet (1974), Goulden and Jackson (1983), Riordan (1958), Stanley (1986), and Stanton and White (1986).

2. Bijections

The method of bijections is really nothing more than the definition of cardinality: two sets have the same number of elements if there is a bijection from one to the other. Thus if we find a bijection between two sets, we have a proof that their cardinalities are equal; and conversely, if we know that two sets have the same cardinality, we may hope to find an explanation in the existence of an easily describable bijection between them. For example, it is very easy to construct bijections between the following three sets: the set of 0–1 sequences of length n, the set of subsets of $[n] = \{1, 2, \ldots, n\}$, and the set of compositions of $n+1$. (A *composition* of an integer is an expression of that integer as sum of positive integers. For example, the compositions of 3 are $1+1+1$, $1+2$, $2+1$, and 3.) The composition $a_1 + a_2 + \cdots + a_k$ of $n+1$ corresponds to the subset $S = \{a_1, a_1 + a_2, \ldots, a_1 + a_2 + \cdots + a_{k-1}\}$ of $\{1, 2, \ldots, n\}$ and to the 0–1 sequence $u_1 u_2 \cdots u_n$ in which $u_i = 1$ if and only if $i \in S$. Moreover, in our example, a composition with k parts corresponds to a subset of cardinality $k-1$ ones, and to a 0–1 sequence with $k-1$ ones, and thus there are $\binom{n}{k-1}$ of each of these.

It is easy to give a bijective proof that the set of compositions of n with parts 1 and 2 is equinumerous with the set of compositions of $n+2$ with all parts at least

2: given a composition $a_1 + a_2 + \cdots + a_k$ of $n+2$ with all $a_i \geqslant 2$, we replace each a_i with

$$2 + \underbrace{1 + \cdots + 1}_{a_i - 2}$$

and then we remove the initial 2. If we let f_n be the number of compositions of n with parts 1 and 2, then f_n is easily seen to satisfy the recurrence $f_n = f_{n-1} + f_{n-2}$ for $n \geqslant 2$, with the initial conditions $f_0 = 1$ and $f_1 = 1$. Thus f_n is a Fibonacci number. (The Fibonacci numbers are usually normalized by $F_0 = 0$ and $F_1 = 1$, so $f_n = F_{n+1}$.)

As another example, if π is a permutation of $[n]$, then we can express π as a product of cycles, where each cycle is of the form $(i \quad \pi(i) \quad \pi^2(i) \quad \cdots \quad \pi^s(i))$. We can also express π as the linear arrangement of $[n]$, $\pi(1)\,\pi(2) \cdots \pi(n)$. Thus the set of cycles $\{(1\ 4), (2), (3\ 5)\}$ corresponds to the linear arrangement 42513. So we have a bijection between sets of cycles and linear arrangements.

This simple bijection turns out to be useful. We use it to give a proof, due to Joyal (1981, p. 16), of Cayley's formula for labeled trees. First note that the bijection implies that for any finite set S the number of sets of cycles of elements of S (each element appearing exactly once in some cycle) is equal to the number of linear arrangements of elements of S.

The number of functions from $[n]$ to $[n]$ is clearly n^n. To each such function f we may associate its *functional digraph* which has an arc from i to $f(i)$ for each i in $[n]$. Now every weakly connected component of a functional digraph (i.e., connected component of the underlying undirected graph) can be represented by a cycle of rooted trees. So by the correspondence just given, n^n is also the number of linear arrangements of rooted trees on $[n]$. We claim now that $n^n = n^2 t_n$, where t_n is the number of trees on $[n]$.

It is clear that $n^2 t_n$ is the number of triples (x, y, T), where $x, y \in [n]$ and T is a tree on $[n]$. Given such a triple, we obtain a linear arrangement of rooted trees by removing all arcs on the unique path from x to y and taking the nodes on this path to be the roots of the trees that remain. This correspondence is bijective, and thus $t_n = n^{n-2}$.

Prüfer (1918) gave a different bijection for Cayley's formula, which is easier to describe but harder to justify. Given a labeled tree on $[n]$, let i_1 be the least leaf (node of degree 1), and suppose that i_1 is adjacent to j_1. Now remove i_1 from the tree and let i_2 be the least leaf of the new tree, and suppose that i_2 is adjacent to j_2. Repeat this procedure until only two nodes are left. Then the original tree is uniquely determined by $j_1 \cdots j_{n-2}$ and conversely any sequence $j_1 \cdots j_{n-2}$ of elements of $[n]$ is obtained from some tree. Thus the number of trees is n^{n-2}.

Both proofs of Cayley's formula sketched above can be refined to count trees according to the number of nodes of each degree, and thereby to prove the Lagrange inversion formula, which we shall discuss in section 6. (See Labelle 1981.)

There is another useful bijection between sets of cycles and linear arrangements which we shall call *Foata's transformation* (see, e.g., Foata 1983) that has interesting properties. Given a permutation in cycle notation, we write each cycle with its least element first, and then we arrange the cycles in decreasing order by their least elements. Thus in our example above, we would have $\pi = (35)(2)(14)$. Then we remove the parentheses to obtain a new permutation whose 1-line notation is $\hat{\pi} = 35214$.

If σ is a permutation of $[n]$, then a *left–right minimum* (or *lower record*) of σ is an index i such that $\sigma(i) < \sigma(j)$ for all $j < i$. It is clear that i is a left–right minimum of $\hat{\pi}$ if and only if $\hat{\pi}(i)$ is the least element in its cycle in π. Thus we have the following.

Theorem 2.1. *The number of permutations of $[n]$ with k left–right minima is equal to the number of permutations of $[n]$ with k cycles.*

This number is (up to sign) a *Stirling number of the first kind*. We shall see them again in sections 3 and 9.

In section 10 we shall need a variant of Foata's transformation in which left–right *maxima* are used instead of left–right minima.

3. Generating functions

The basic idea of generating functions is the following: instead of finding the cardinality of a set S, we assign to each α in S a *weight* $w(\alpha)$. Then the *generating function* $\mathscr{G}(S)$ for S (with respect to the weighting function w) is $\sum_{\alpha \in S} w(\alpha)$. Thus the concept of generating function for a set is a generalization of the concept of cardinality. Note that S may be infinite as long as the sum converges (often as a formal power series).

The weights may be elements of any abelian group, but they are usually monomials in a ring of polynomials or power series. In a typical application each element α of S will have a "length" $l(\alpha)$ and we take the weight of α to be $x^{l(\alpha)}$, where x is an indeterminate. Then knowing the generating function $\sum_{\alpha \in S} x^{l(\alpha)}$ is equivalent to knowing the number of elements of S of each length.

Analogous to the product rule for cardinalities, $|A||B| = |A \times B|$, is the product rule for generating functions, $\mathscr{G}(A)\mathscr{G}(B) = \mathscr{G}(A \times B)$, where we take the "product weight" on $A \times B$, defined by $w\big((\alpha, \beta)\big) = w(\alpha)w(\beta)$.

As an example, suppose we want to count sequences of zeros and ones of length n according to the number of zeros they contain. We can identify the set of 0–1 sequences of length n with the Cartesian product $\{0,1\}^n$. If we weight $\{0,1\}$ by $w(0) = x$ and $w(1) = y$, then the product weight on $\{0,1\}^n$ assigns to a sequence with j zeros and $k\ (= n - j)$ ones the weight $x^j y^k$. Thus

$$(x+y)^n = \sum_{j+k=n} \binom{n}{j} x^j y^k$$

is the generating function for sequences of zeros and ones of length n by the number of zeros and the number of ones. If we want to count sequences of zeros and ones of all lengths with this weighting, we sum on n to obtain the generating function

$$\sum_{j,k=0}^{\infty}\binom{j+k}{j}x^jy^k=\frac{1}{1-x-y}.$$

Now suppose we want to count compositions with parts 1 and 2. Rather than picking an integer n and considering the compositions of n, we pick an integer k and consider the set C_k of all compositions of any integer with exactly k parts (each part being 1 or 2). We may identify C_k with $\{1,2\}^k$. If we assign 1 the weight x and 2 the weight x^2, where x is an indeterminate, then the product weight of a composition of n in C_k is x^n. Thus

$$\mathscr{G}(C_k)=\mathscr{G}(\{1,2\}^k)=\mathscr{G}(\{1,2\})^k=(x+x^2)^k=\sum_{n=k}^{2k}\binom{k}{n-k}x^n.$$

Thus there are $\binom{k}{n-k}$ compositions of n with k parts, each part 1 or 2. As before, if we do not care about the number of parts, we sum on k to obtain $\sum_{k=0}^{\infty}(x+x^2)^k=(1-x-x^2)^{-1}$ as the generating function for all compositions into parts 1 and 2. By the same kind of reasoning, if A is any set of positive integers, then the generating function for compositions with k parts, all in A, is $\left(\sum_{i\in A}x^i\right)^k$ and the generating function for compositions with any number of parts, all in A, is $\left(1-\sum_{i\in A}x^i\right)^{-1}$. In particular, if A is the set of positive integers then $\sum_{i\in A}x^i=x/(1-x)$ so the generating function for compositions with k parts is

$$\left(\frac{x}{1-x}\right)^k=\sum_{n=k}^{\infty}\binom{n-1}{k-1}x^n$$

for $k>0$ and the generating function for all compositions is

$$\left(1-\frac{x}{1-x}\right)^{-1}=\frac{1-x}{1-2x}=1+\frac{x}{1-2x}=1+\sum_{n=1}^{\infty}2^{n-1}x^n.$$

The generating function for compositions with parts greater than 1 is

$$\left(1-\frac{x^2}{1-x}\right)^{-1}=\frac{1-x}{1-x-x^2}=1+\frac{x^2}{1-x-x^2}=1+\sum_{n=1}^{\infty}F_{n-1}x^n.$$

Similarly, the generating function for compositions with odd parts is

$$\left(1-\frac{x}{1-x^2}\right)^{-1}=1+\frac{x}{1-x-x^2}=1+\sum_{n=1}^{\infty}F_nx^n.$$

Thus we have proved using generating functions two of the results we proved using bijections in the preceding section. Notice that the generating functions take into account initial cases which did not arise in the bijective approach.

For our next two examples, we consider another bijection for permutations. Suppose that π is a permutation of $[n]$. We associate to π a sequence $a_1a_2\cdots a_n$ of integers satisfying $0 \leqslant a_j \leqslant j-1$ for each j as follows: a_j is the number of indices $i<j$ for which $\pi(i) > \pi(j)$. The sequence $a_1a_2\cdots a_n$ is called the *inversion table* of π: an inversion of π is a pair (i,j) with $i<j$ and $\pi(i)>\pi(j)$, and thus a_j is the number of inversions of π of the form (i,j). It is not difficult to show that the correspondence between permutations and their inversion tables gives a bijection between the set $\mathcal{S}_n$ of permutations of $[n]$ and the set T_n of sequences $a_1a_2\cdots a_n$ of integers satisfying $0 \leqslant a_j \leqslant j-1$ for each j. Note that T_n is the Cartesian product $\{0\} \times \{0,1\} \times \cdots \times \{0,1,\ldots,n-1\}$. We shall use the inversion table to count permutations by inversions and also by cycles.

Let $I(\pi)$ be the number of inversions of π. We would like to find the generating function $\mathcal{G}(\mathcal{S}_n)$ for permutations of $[n]$ where each permutation π is assigned the weight $q^{I(\pi)}$. To do this we note that $I(\pi)$ is the sum of the entries of the inversion table of π, and thus if we assign the weight $w(a) = q^{a_1+\cdots+a_n}$ to $a = a_1a_2\cdots a_n \in T_n$, then we have

$$\mathcal{G}(\mathcal{S}_n) = \mathcal{G}(T_n) = 1\cdot(1+q)\cdots(1+q+\cdots+q^{n-1}).$$

Next we count permutations by left–right minima. It is clear that j is a left–right minimum of π if and only if $a_j = j-1$. Thus if we assign the weight t^k to a permutation in $\mathcal{S}_n$ with k left–right minima, and to a sequence in T_n with k occurrences of $a_j = j-1$, then we have

$$\mathcal{G}(\mathcal{S}_n) = \mathcal{G}(T_n) = t(t+1)(t+2)\cdots(t+n-1) = \sum_{k=0}^{n} c(n,k)t^k,$$

where $c(n,k)$ is (by definition) the unsigned Stirling number of the first kind. By Theorem 2.1, it follows that $c(n,k)$ is also the number of permutations in $\mathcal{S}_n$ with k cycles.

4. Free monoids

Free monoids provide a useful way of organizing many simple applications of generating functions. Let A be a set of "letters". The *free monoid* A^* is the set of all finite sequences (including the empty sequence) of elements of A, usually called *words*, with the operation of concatenation. We can construct an algebra from A^* by taking formal sums of elements of A^* with coefficients in some ring. We write 1 for the empty sequence, which is the unit of this algebra. These formal sums are then formal power series in noncommuting variables. The generating function $\mathcal{G}(S)$ for a subset S of A^* is the sum of its elements.

If S and T are subsets of A^*, we write ST for the set $\{st \mid s \in S \text{ and } t \in T\}$. We say that the product ST is *unique* if every element of ST has only one such factorization. The fundamental fact about generating functions is that if the product ST is unique, then $\mathscr{G}(ST) = \mathscr{G}(S)\mathscr{G}(T)$.

More generally, we may define a free monoid to be a set together with an associative binary operation which is isomorphic to a free monoid as defined above. Let $A^+ = A^* \setminus \{1\}$ and suppose that S is a subset of A^+ such that for each k, every element of S^k has a unique factorization $s_1 s_2 \cdots s_k$ with each s_i in S. Such a set S is sometimes called a *uniquely decodable code*, or simply a *code*. Then $S^* = \bigcup_{i=0}^{\infty} S^i$ is a free monoid. We call the elements of S the *primes* of the free monoid S^*. In this case

$$\mathscr{G}(S^*) = \sum_{i=0}^{\infty} \mathscr{G}(S)^i = \big(1 - \mathscr{G}(S)\big)^{-1}.$$

In particular, $\mathscr{G}(A^*) = \big(1 - \mathscr{G}(A)\big)^{-1}$.

Among the simplest free monoid problems are those dealing with compositions of integers, as we saw in the previous section. A composition of an integer is simply an element of the free monoid $\mathbb{P}^*$, where $\mathbb{P}$ is the set of positive integers.

As a more interesting example, let $A = \{X, Y\}$, let S be the subset of A^* consisting of words with equal numbers of X's and Y's, and let T be the subset of A^* of words with no nonempty initial segment in S. Then $A^* = ST$ uniquely, so $\mathscr{G}(A^*) = (1 - X - Y)^{-1} = \mathscr{G}(S)\mathscr{G}(T)$. Moreover, S is a free monoid U^*, where U is the set of words in S which cannot be factored nontrivially in S. The sets S, T, and U have simple interpretations in terms of walks in the plane, starting at the origin. If X and Y are represented by unit steps in the x and y directions, then S corresponds to walks which end on the main diagonal, T corresponds to walks that never return to the main diagonal, and U corresponds to walks that return to the main diagonal only at the end.

It is often useful to replace the noncommuting variables by commuting variables. If we replace the letter X by the variable x, we are *assigning X the weight x*. (More formally, we are applying a homomorphism in which the image of X is x.)

In our example, if we weight X and Y by commuting variables x and y, then $\mathscr{G}(A^*)$ becomes $1/(1-x-y)$ and $\mathscr{G}(S)$ becomes $\sum_{n=0}^{\infty} \binom{2n}{n} x^n y^n = (1-4xy)^{-1/2}$ since there are $\binom{2n}{n}$ ways of arranging n X's and n Y's. Thus $\mathscr{G}(T)$ becomes $\sqrt{1-4xy}/(1-x-y)$. It can be shown that this is equal to

$$\sum_{m,n=0}^{\infty} \frac{|m-n|}{m+n} \binom{m+n}{n} x^m y^n, \tag{4.1}$$

where the constant term is 1. The coefficients in (4.1) are called *ballot numbers* and we shall see them again in section 6.

If we replace x and y by the same variable z, the generating function for T becomes

$$\frac{\sqrt{1-4z^2}}{1-2z} = \frac{1+2z}{\sqrt{1-4z^2}}.$$

Thus the number of words in T of length $2n$ is $\binom{2n}{n}$ and the number of words in T of length $2n+1$ is $2\binom{2n}{n}$.

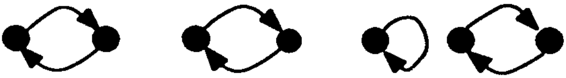
Although we usually work with formal power series, it is sometimes useful for variables to take on real values. We derive an inequality called *McMillan's inequality* which is useful in information theory. (See McMillan 1956.) Let A be an alphabet (set of letters) of size r, and let S be a code in A^*, so that S^* is a free monoid. Let us weight each letter of A by t, and let $\mathscr{G}(S) = p(t)$.

We know that $\mathscr{G}(S^*) = (1-p(t))^{-1}$ as formal power series in t. Since there are r^k words in A^* of length k, the coefficient of t^k in $(1-p(t))^{-1}$ is at most r^k. If $0 < \alpha < 1/r$ then the series $\sum_{k=0}^{\infty} r^k \alpha^k$ converges absolutely to $(1-r\alpha)^{-1}$, and thus $(1-p(\alpha))^{-1} \leqslant (1-r\alpha)^{-1}$, which implies $p(\alpha) \leqslant r\alpha$. Taking the limit as α approaches $1/r$ from below, we obtain $p(1/r) \leqslant 1$.

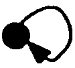
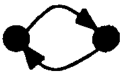
Thus we have proved the following.

Theorem 4.2. *Let S be a uniquely decodable code in an alphabet of size r, and for each k let p_k be the number of words in S of length k. Then $\sum_{k=1}^{\infty} p_k r^{-k} \leqslant 1$.*

In some applications of free monoids, the "letters" have some internal structure. For example, consider the set of permutations π of $[n] = \{1, 2, \ldots n\}$ satisfying $|\pi(i) - i| \leqslant 1$. We can represent a permutation of $[n]$ as a digraph with node set $[n]$ with an arc from i to $\pi(i)$ for each i. If we draw the digraph with the nodes in increasing order, we get a picture like this one, which corresponds to the permutation 2 1 4 3 5 7 6:

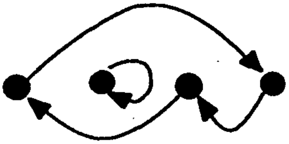


It is clear that these permutations form a free monoid with the two "letters", or primes

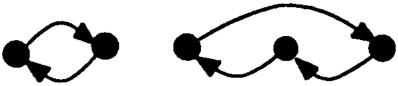
 and 

Thus the generating function (by length) for these permutations is $(1-x-x^2)^{-1}$, and there is an obvious bijection between these permutations and compositions with parts 1 and 2.

Sometimes it is easier to count all the elements of a free monoid than just the primes. If we represent arbitrary permutations as in the previous example, then we have a free monoid in which the primes, called *indecomposable permutations*, are those permutations π of $[n]$ (for some n) such that for $1 \leqslant i < n$, π restricted to $[i]$ is not a permutation of $[i]$. For example, 4213 is indecomposable:



but 21534 is not:



Thus if $g(x)$ is the generating function for indecomposable permutations, we have

$$\sum_{n=0}^{\infty} n!\,x^n = (1-g(x))^{-1},$$

so

$$g(x) = 1 - \left(\sum_{n=0}^{\infty} n!\,x^n\right)^{-1},$$

as shown by Comtet (1972).

5. Circular words

We now study some properties of words in which the letters are thought of as arranged in a circle, so that the last letter is considered to be followed by the first. (This should not be confused with the problem of counting equivalence classes of words under cyclic permutation, which we discuss in section 14.)

We define the *cyclic shift operator* C on words by

$$C\,a_1a_2\cdots a_k = a_2a_3\cdots a_ka_1.$$

A *conjugate* or *cyclic permutation* of a word w is a word of the form $C^m w$ for some m. If S is a set of words, then we define S° to be the set of all conjugates of words in S.

Suppose that S^* is a free submonoid of the free monoid A^*, and let $w = s_1s_2\cdots s_k$ be an element of S^k, where each s_i is in S. It is clear that $C^iw \in S^k$ whenever i takes on any of the k values $0, l(s_1), l(s_1s_2), \ldots, l(s_1s_2\cdots s_{k-1})$, where $l(v)$ denotes the length of the word v. If these are the only values of i, with $0 \leqslant i < l(w)$, for which $C^iw \in S^*$, then we call S^* *cyclically free*[1]. For example, $\{ab, b\}^*$ is cyclically free, but $\{aa\}^*$ is not.

If S^* is cyclically free, then it is clear that for $w \in (S^k)^\circ$ there are exactly k values of i, with $0 \leqslant i < l(w)$, for which $C^iw \in S^k$.

Theorem 5.1. *Suppose that S^* is cyclically free and let $Q = S^k \cap A^n$. Then $k|Q^\circ| = n|Q|$.*

Proof. We count pairs (i, w), where $C^iw \in Q$ and $0 \leqslant i < n$. First we may choose C^iw in $|Q|$ ways. Then w is determined by i, which may be chosen arbitrarily in $\{0, 1, \cdots, n-1\}$. Thus there are $n|Q|$ pairs. On the other hand, we may choose w first as an arbitrary element of Q° and by the remark above, there are k choices for i. □

In the next section we shall use a weighted version of Theorem 5.1 which is proved exactly the same way.

From Theorem 5.1 we can easily derive a generating function for $(S^*)^\circ$:

Corollary 5.2. *Suppose that S^* is cyclically free and let $g(z) = \sum_{w\in S} z^{l(w)}$. Then*

$$\sum_{n,k=1}^{\infty} |(S^k)^\circ \cap A^n| t^k \frac{z^n}{n} = \log \frac{1}{1 - tg(z)}.$$

Equivalently,

$$\sum_{n,k=1}^{\infty} |(S^k)^\circ \cap A^n| t^k z^n = \frac{tzg'(z)}{1 - tg(z)}.$$

We can use Theorem 5.1 to count the number of k-subsets of $[n]$ with no two consecutive elements, where 1 and n are considered consecutive. We take $S = \{ab, b\}$. The subsets we want correspond to words in $(S^{n-k} \cap \{a, b\}^n)^\circ$. These words contain n letters, of which $n-k$ are b's, and hence k are a's. The positions of the a's in one of these words determines the subset. S^* is clearly cyclically free, so by Theorem 5.1, the number of such subsets is $(n/(n-k))\binom{n-k}{k}$.

Our next example will be useful in proving the Lagrange inversion formula in the next section. Let ϕ be any function from A to the real numbers. Extend ϕ to all of A^* by defining $\phi(a_1a_2\cdots a_k) = \phi(a_1) + \cdots + \phi(a_k)$. Define R by

$$R = \{ w \mid \text{if } w = uv \text{ with } u \neq 1 \text{ then } \phi(u) < 0 \}. \tag{5.3}$$

[1] In the theory of codes, S is called a *circular code* and S^* is called a *very pure* free monoid. See, e.g., Berstel and Perrin (1983).

It is easily verified that R is a cyclically free submonoid of A^*.

The following description of R° is the key step in our proof of the Lagrange inversion formula in the next section.

Lemma 5.4. *Let R and ϕ be as above. Then $R^\circ = \{1\} \cup \{w \mid \phi(w) < 0\}$.*

Proof. We need only show that if $\phi(w) < 0$, then for some i, $C^i w \in R$. Of the heads (initial segments) h of w which maximize $\phi(h)$, let u be the longest, and let $w = uv$. Then vu is easily verified to be in R. □

6. Lagrange inversion

In the last example, let $A = \{x_{-1}, x_0, x_1, x_2, \ldots\}$ and define $\phi : A^* \to \mathbb{Z}$ by

$$\phi(x_{i_1} \cdots x_{i_m}) = i_1 + \cdots + i_m.$$

Let R be as in (5.3) and let $S = \{w \mid w \in R \text{ and } \phi(w) = -1.\}$. We claim that $R = S^*$. Since we know that R is a free monoid, we need only show that if w is a prime of R, then $\phi(w) = -1$.

To see this, let w be a prime of R. Since $\phi(w) < 0$ and $\phi(x_i) \geqslant -1$ for each $x_i \in A$, w must have a head h with $\phi(h) = -1$. Let $w = uv$, where u is the longest head of w for which $\phi(u) = -1$. Then v must be in R, since otherwise w would have a longer head h with $\phi(h) \geqslant -1$. Since w is a prime of R, this means $w = u$. It follows that if v is any word in R with $\phi(v) = -k$ then $v \in S^k$.

Now let v be any word in S and suppose $v = ux_i$. Then u is in R with $\phi(u) + i = -1$, so $\phi(u) = -1 - i$, and thus $u \in S^{i+1}$. It follows that

$$S = \bigcup_{i=-1}^{\infty} S^{i+1} x_i, \tag{6.1}$$

where the union is disjoint. We are now ready to prove the Lagrange inversion formula. We use the notation $[x^n]F(x)$ to denote the coefficient of x^n in $F(x)$.

Theorem 6.2. *Let $g(u) = \sum_{n=0}^{\infty} g_n u^n$, where the g_n are indeterminates. Then there is a unique formal power series f in the g_n satisfying $f = g(f)$, and for $k > 0$,*

$$f^k = \sum_{n=1}^{\infty} \frac{k}{n} [u^{n-k}] g(u)^n. \tag{6.3}$$

Proof. It is easily seen that the equation $f = g(f)$ has a unique solution. Let us assign to the letter x_i the weight g_{i+1} and let f be the image of $\mathcal{G}(S)$ under this assignment. Then from (6.1) we have

$$f = \sum_{i=-1}^{\infty} f^{i+1} g_{i+1} = g(f).$$

By the weighted version of Theorem 5.1, the sum of the weights of the words of length n in S^k is k/n times the sum of the weights of the words in $(S^k \cap A^n)^\circ$. But by Lemma 5.4, the sum of the weights of the words in $(S^k \cap A^n)^\circ$ is

$$[u^{-k}]\left(\frac{g(u)}{u}\right)^n = [u^{n-k}]g(u)^n.$$

The proof we have just given is essentially that of Raney (1960). It is clear that if the g_i are assigned values that are not necessarily indeterminates, then the theorem still holds as long as the sum in (6.3) converges as a formal power series and f is uniquely determined as a formal power series by $f = g(f)$. The usual formulation of Lagrange inversion is obtained by taking $g(u) = z\sum_{n=0}^{\infty} r_n u^n$, where z is an indeterminate and the r_n are arbitrary.

One of the most important applications of Lagrange inversion is to the enumeration of ordered trees. (An ordered tree is a rooted unlabeled tree in which the children of any node are linearly ordered.) Let us weight a node with i children in an ordered tree by g_i, and weight the tree by the product of the weights of its nodes. If f is the sum of the weights of all ordered trees, then, since an ordered tree consists of a root together with some number (possibly zero) of children, each of which may be an arbitrary ordered tree, we have

$$f = \sum_{i=0}^{\infty} g_i f^i = g(f),$$

where $g(u) = \sum_{i=0}^{\infty} g_i u^i$. The Lagrange inversion formula then yields the following.

Theorem 6.4. *The number of k-tuples of ordered trees in which a total of n_i nodes have i children is*

$$\frac{k}{n}\binom{n}{n_0, n_1, n_2, \ldots}, \quad \textit{where } n = \textstyle\sum_i n_i,$$

if $n_1 + 2n_2 + 3n_3 + \cdots = n - k$, and 0 otherwise.

It is not hard to derive Theorem 6.2 from Theorem 6.4, so any other proof of Theorem 6.4 (for example, by induction), yields a proof of the Lagrange inversion formula. Our approach can also be used to give a purely combinatorial proof of Theorem 6.4 without the use of generating functions.

A few special cases of Theorems 6.2 and 6.4 are especially important. If there are a nodes with 2 children, b nodes with no children, and no other nodes, then with $b = a + k$ the number of k-tuples of such trees is

$$\frac{k}{n}\binom{n}{a} = \frac{k}{2a+k}\binom{2a+k}{a}.$$

These numbers are called *ballot numbers*. The special case $k = 1$ gives the *Catalan numbers*

$$\frac{1}{2a+1}\binom{2a+1}{a} = \frac{1}{a+1}\binom{2a}{a}.$$

To apply Theorem 6.2 directly to this case, we may take $g_0 = 1$, $g_2 = x$, and $g_i = 0$ for $i \neq 0,2$. Then f satisfies $f = 1 + xf^2$, so $f = \left(1 - \sqrt{1-4x}\right)/2x$, and we obtain

$$\left(\frac{1-\sqrt{1-4x}}{2x}\right)^k = \sum_{a=0}^{\infty} \frac{k}{2a+k}\binom{2a+k}{a}x^a.$$

To count all ordered trees we set $g_i = x$ for all i, to obtain the equation $f(x) = x/\left(1 - f(x)\right)$, with the solution

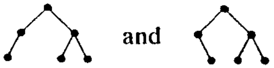
$$\begin{aligned} f(x)^k &= \left(\frac{1-\sqrt{1-4x}}{2}\right)^k \\ &= \sum_{n=0}^{\infty} \frac{k}{n}\binom{2n-k-1}{n-k}x^n = \sum_{n=0}^{\infty} \frac{k}{2n+k}\binom{2n+k}{n}x^{n+k}, \end{aligned}$$

so we again obtain the Catalan and ballot numbers. It is an instructive exercise to find a bijection between these classes of trees, and to relate these results to formula (4.1).

Our analysis gives a well-known bijection between ordered trees and words in S. The code $\mathrm{c}(t)$ for a tree T may be defined as follows: If the root of T has no children, then $\mathrm{c}(T) = x_{-1}$. Otherwise, if the children of the root of T are (in order) the roots of trees $T_1, T_2, \ldots, T_k$, then

$$\mathrm{c}(T) = \mathrm{c}(T_1)\cdots\mathrm{c}(T_k)x_{k-1}.$$

For another example, we define a *binary tree* to be a rooted tree in which every node has a left child, a right child, neither, or both. Thus



are different binary trees. Let us weight a binary tree with n nodes, i left children, and j right children by $x^nL^iR^j$. Then if f is the generating function for these trees, we have

$$f = x(1 + Lf)(1 + Rf),$$

and thus by Lagrange inversion we have

$$f^k = \sum_{n=k}^{\infty}\sum_{i=0}^{n-k} \frac{k}{n}\binom{n}{i}\binom{n}{i+k}L^iR^{n-k-i}x^n.$$

For $k = 1$, the numbers $(1/n)\binom{n}{i}\binom{n}{i+1}$ are called *Runyon numbers* or *Narayana numbers*.

7. The transfer matrix method

Many enumeration problems can be transformed into problems of counting walks in digraphs, which can be solved by the transfer matrix method. Suppose D is a finite digraph. To every arc of D we associate a weight. Let M be the matrix in which rows and columns are indexed by the nodes of D and the (i,j) entry of M is the sum of the weights of the arcs from i to j. Then by the definition of matrix multiplication, the (i,j) entry in M^k is the sum of the weights of all walks of k arcs from i to j. It follows that (as long as the infinite sum exists) $\sum_{k=0}^{\infty} M^k = (I-M)^{-1}$ counts all walks, where I is the identity matrix, and trace $(I-M)^{-1}$ counts walks that end where they begin.

For example, consider the following problem: Given integers n and i, what is the number $t(n,i)$ of sequences $a_1a_2\cdots a_n$ of 0's, 1's, and -1's with $a_1+\cdots+a_n \equiv i \pmod 6$? Here we take D to be the digraph with node set $\{0,1,2,3,4,5\}$ and an arc from each j to $j-1$, j, and $j+1$, reduced modulo 6. We weight each arc by x. So M is

$$\begin{pmatrix} x & x & 0 & 0 & 0 & x \\ x & x & x & 0 & 0 & 0 \\ 0 & x & x & x & 0 & 0 \\ 0 & 0 & x & x & x & 0 \\ 0 & 0 & 0 & x & x & x \\ x & 0 & 0 & 0 & x & x \end{pmatrix}.$$

We find that $(I-M)^{-1}$ is the circulant matrix with first column

$$\frac{1}{6}\begin{pmatrix} \frac{1}{1-3x}+\frac{2}{1-2x}+\frac{1}{1+x}+2 \\ \frac{1}{1-3x}+\frac{1}{1-2x}-\frac{1}{1+x}-1 \\ \frac{1}{1-3x}-\frac{1}{1-2x}+\frac{1}{1+x}-1 \\ \frac{1}{1-3x}-\frac{2}{1-2x}-\frac{1}{1+x}+2 \\ \frac{1}{1-3x}-\frac{1}{1-2x}+\frac{1}{1+x}-1 \\ \frac{1}{1-3x}+\frac{1}{1-2x}-\frac{1}{1+x}-1 \end{pmatrix}.$$

Thus for $n>0$,

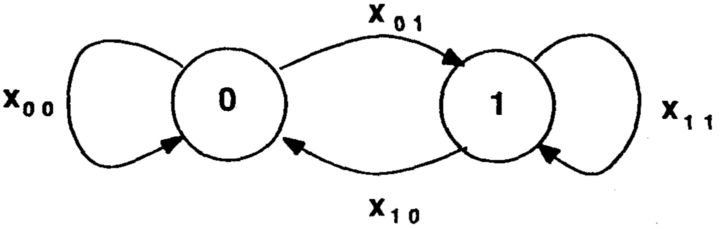
$$t(n,0) = (3^n + 2^{n+1} + (-1)^n)/6,$$

$$t(n,1) = t(n,5) = (3^n + 2^n - (-1)^n)/6,$$

$$t(n,2) = t(n,4) = (3^n - 2^n + (-1)^n)/6,$$

$$t(n,3) = (3^n - 2^{n+1} - (-1)^n)/6.$$

As another example, how many 0–1 sequences are there with specified numbers of occurrences of 00, 01, 10, and 11? Here we take D to be the weighted digraph



Then

$$(I-M)^{-1} = \begin{pmatrix} 1-x_{00} & -x_{01} \\ -x_{10} & 1-x_{11} \end{pmatrix}^{-1}$$
$$= \frac{\begin{pmatrix} 1-x_{11} & x_{01} \\ x_{10} & 1-x_{00} \end{pmatrix}}{(1-x_{00})(1-x_{11})-x_{01}x_{10}}.$$

Thus, for example, the generating function for 0–1 sequences beginning with 0 and ending with 1 is

$$\frac{x_{01}}{(1-x_{00})(1-x_{11})-x_{01}x_{10}} = \sum_{i=0}^{\infty} \frac{x_{01}^{i+1}x_{10}^{i}}{(1-x_{00})^{i+1}(1-x_{11})^{i+1}}$$
$$= \sum_{i,j,k} x_{01}^{i+1}x_{10}^{i}x_{00}^{j}x_{11}^{k}\binom{i+j}{j}\binom{i+k}{k},$$

so $\binom{i+j}{j}\binom{i+k}{k}$ is the number of 0–1 sequences beginning with 0 and ending with 1, with i occurrences of 10 (and thus $i+1$ occurrences of 01), j occurrences of 00, and k occurrences of 11 (and thus $i+j+1$ zeros and $i+k+1$ ones).

The transfer matrix method can often be used to show that a generating function is rational. For example, consider the problem of counting the number of ways of covering an $m \times n$ rectangle with a fixed finite set of polyominos. It is not hard to show, using the transfer matrix method, that for fixed m the generating function on n is rational, although it is difficult to give an explicit formula. We will see another example of this type in section 10.

8. Multisets and partitions

We have so far considered problems involving linear arrangements. In this and the next section we turn to unordered collections. We first consider the problem of counting *multisets*, which are sets with repeated elements allowed. More formally, a multiset on a set S is a function from S to the nonnegative integers; if ν is a multiset then $\nu(s)$ represents the multiplicity of s. If each element s in S has a weight $w(s)$, then we define the weight of the multiset ν to be $\prod_{s\in S} w(s)^{\nu(s)}$.

For each s in S, let M_s be a set of nonnegative integers. Then the sum of the weights of all multisets ν on S such that $\nu(s)$ is in M_s for each s in S is easily seen to be

$$\prod_{s\in S}\sum_{i\in M_s} w(s)^i.$$

We give a few examples. Let us take $w(s)=x$ for all s in S, and assume $|S|=n$. If $M_s=\{0,1\}$ for each s, we are counting subsets, and the generating function is

$$(1+x)^n=\sum_{k=0}^{n}\binom{n}{k}x^k.$$

If M_s is the set of all nonnegative integers for each s, we are counting unrestricted multisets, and the generating function is

$$(1+x+x^2+\cdots)^n=(1-x)^{-n}=\sum_{k=0}^{\infty}\binom{n+k-1}{k}x^k.$$

If $M_s=\{0,1,\ldots,m\}$ for each s, the generating function is

$$\begin{aligned}(1+x+\cdots+x^m)^n&=\left(\frac{1-x^{m+1}}{1-x}\right)^n\\&=\sum_{k=0}^{\infty}x^k\sum_i(-1)^i\binom{n}{i}\binom{n+k-(m+1)i-1}{k-(m+1)i}.\end{aligned}$$

A multiset of positive integers with sum k is called a *partition* of k. The elements of a partition are called its *parts*. It is customary to list the parts of a partition in decreasing order, so a partition of k is often defined as a (weakly) decreasing sequence of positive integers with sum k. To count partitions, we weight i by q^i, where q is an indeterminate. Then the generating function for all partitions is $\prod_{i=1}^{\infty}(1-q^i)^{-1}$ and the generating function for partitions with distinct parts is $\prod_{i=1}^{\infty}(1+q^i)$.

Many theorems in the theory of partitions assert that one set of partitions is equinumerous with another. The simplest of these, due to Euler, is that the number of partitions of n with odd parts is equal to the number of partitions of n with distinct parts. To prove this, we note that the generating function for partitions with odd parts is

$$\begin{aligned}\prod_{i\text{ odd}}(1-q^i)^{-1}&=\prod_{i=1}^{\infty}(1-q^i)^{-1}\prod_{j=1}^{\infty}(1-q^{2j})\\&=\prod_{i=1}^{\infty}\frac{1-q^{2i}}{1-q^i}=\prod_{i=1}^{\infty}(1+q^i),\end{aligned}$$

which is the generating function for partitions with distinct parts.

It is not difficult to give a combinatorial proof of this result: Suppose π is a partition with odd parts. If π contains the odd part i with multiplicity k, let $k = 2^{e_1} + 2^{e_2} + \cdots + 2^{e_s}$, where $0 \leqslant e_1 < e_2 < \cdots < e_s$. We now replace the k copies of part i by the distinct parts $2^{e_1}i, 2^{e_2}i, \ldots, 2^{e_s}i$. Doing this to every part of π we obtain a partition π' with distinct parts. The correspondence is easily seen to be a bijection. For example, if $\pi = \{9,9,7,7,7,1,1,1,1\}$, then $\pi' = \{18,14,7,4\}$.

One of the most famous results in the theory of partitions is the following.

Theorem 8.1. *The number of partitions of n with distinct parts in which any two parts differ by at least 2 is equal to the number of partitions of n with parts congruent to* 1 *or* 4 (mod 5).

This result follows easily from the *Rogers–Ramanujan identity*

$$\sum_{i=0}^{\infty} \frac{q^{i^2}}{(1-q)(1-q^2)\cdots(1-q^i)} = \prod_{j=0}^{\infty} \frac{1}{(1-q^{5j+1})(1-q^{5j+4})}.$$

No simple bijective proof of Theorem 8.1 is known. A complicated bijective proof was found by Garsia and Milne (1981).

We now prove an identity called the *q-binomial theorem*, which has many applications to partitions. We introduce the notation $(a)_n$ for $(1-a)(1-aq)\cdots(1-aq^{n-1})$, where q is understood. In particular, $(q)_n = (1-q)(1-q^2)\cdots(1-q^n)$. We also write $(a)_\infty$ for $\prod_{i=0}^{\infty}(1-aq^i)$.

Theorem 8.2 (The q-binomial theorem).

$$\sum_{n=0}^{\infty} \frac{(a)_n}{(q)_n} t^n = \frac{(at)_\infty}{(t)_\infty}.$$

Proof. Let

$$\frac{(at)_\infty}{(t)_\infty} = \sum_{n=0}^{\infty} f_n t^n.$$

Then

$$\frac{(at)_\infty}{(tq)_\infty} = (1-t)\frac{(at)_\infty}{(t)_\infty} = (1-t)\sum_{n=0}^{\infty} f_n t^n.$$

But also,

$$\frac{(at)_\infty}{(tq)_\infty} = (1-at)\frac{(atq)_\infty}{(tq)_\infty} = (1-at)\sum_{n=0}^{\infty} f_n q^n t^n.$$

Equating coefficients of t^n, we have

$$f_n - f_{n-1} = q^n f_n - aq^{n-1} f_{n-1}, \quad n \geqslant 1,$$

and thus $f_n(1-q^n) = f_{n-1}(1-aq^{n-1})$. Since $f_0 = 1$, this gives

$$f_n = \prod_{i=1}^{n} \frac{1-aq^{i-1}}{1-q^i} = \frac{(a)_n}{(q)_n}. \qquad \square$$

Two cases are particularly worth noting. If $a = q^m$, where m is a positive integer, then we have

$$\sum_{n=0}^{\infty} \frac{(q^m)_n}{(q)_n} t^n = \frac{1}{(t)_m}. \tag{8.3}$$

The *q-binomial coefficient* is defined to be

$$\begin{bmatrix} n \\ k \end{bmatrix} = \frac{(q)_n}{(q)_k (q)_{n-k}}.$$

Since $(q^m)_n = (q)_{m+n-1}/(q)_{m-1}$, we may rewrite (8.3) as

$$\sum_{n=0}^{\infty} \begin{bmatrix} m+n-1 \\ n \end{bmatrix} t^n = \frac{1}{(1-t)(1-tq)\cdots(1-tq^{m-1})}. \tag{8.4}$$

It follows from (8.4) that $\left[\begin{smallmatrix} n \\ k \end{smallmatrix}\right]$ is a polynomial in q that reduces to the binomial coefficient $\binom{n}{k}$ for $q = 1$.

We can use (8.4) to count partitions with at most n parts, each part at most m. It is clear that the desired generating function is the coefficient of t^n in

$$\frac{1}{(1-t)(1-tq)\cdots(1-tq^m)} = \frac{1}{(t)_{m+1}},$$

and by (8.4) this is $\left[\begin{smallmatrix} m+n \\ n \end{smallmatrix}\right]$.

The case $a = q^{-m}$ of the q-binomial theorem yields similarly (after changing q to q^{-1} and t to $-t/q$)

$$\sum_{n=0}^{m} t^n q^{\binom{n}{2}} \begin{bmatrix} m \\ n \end{bmatrix} = (1+t)(1+tq)\cdots(1+tq^{m-1}), \tag{8.5}$$

which implies that the generating function for partitions with n distinct parts, all less than m, where 0 is allowed as a part, is $q^{\binom{n}{2}}\left[\begin{smallmatrix} m \\ n \end{smallmatrix}\right]$. This result may be derived directly from our previous generating function for partitions with repeated parts allowed, since every partition with distinct parts is obtained uniquely from an unrestricted partition by adding 0 to the smallest part, 1 to the next smallest, and so on.

There is an important interpretation for q-binomial coefficients in terms of vector spaces over finite fields. (See, e.g., Stanley 1986, p. 28, for the proof.)

Theorem 8.6. *Let q be a prime power. Then the number of k-dimensional subspaces of an n-dimensional vector space over a field with q elements is $\left[\begin{smallmatrix} n \\ k \end{smallmatrix}\right]$.*

A comprehensive reference on the theory of partitions is Andrews (1976).

9. Exponential generating functions

If $a_0, a_1, \ldots$ is a sequence of numbers, the power series

$$\sum_{n=0}^{\infty} a_n \frac{x^n}{n!}$$

is called the *exponential generating function* for the sequence. Exponential generating functions arise in counting "labeled objects". Their usefulness comes from the fact that

$$\frac{x^m}{m!}\frac{x^n}{n!} = \binom{m+n}{m}\frac{x^{m+n}}{(m+n)!}.$$

If A is an object with label set $[m]$ and B is an object with label set $[n]$, we can combine them in $\binom{m+n}{m}$ ways to get an object (A', B') with label set $[m+n]$: We first choose an m-element subset S of $[m+n]$ and replace the labels of A with the elements of S (preserving their order) to get A', and in the same way we get B' from B and $[m+n] \setminus S$.

Thus if $f(x)$ and $g(x)$ are exponential generating functions for classes of labeled objects, then their product $f(x)g(x)$ will be the exponential generating function for ordered pairs of these objects. For example, the exponential generating function for nonempty sets is

$$e^x - 1 = \sum_{n=1}^{\infty} \frac{x^n}{n!}$$

since the elements of $[n]$ can be arranged as a nonempty set in one way if $n > 0$ and in zero ways if $n = 0$. Thus

$$(e^x - 1)^2 = \sum_{n=2}^{\infty} (2^n - 2)\frac{x^n}{n!}$$

is the exponential generating function for ordered partitions of a set into two nonempty blocks. More generally, $(e^x - 1)^k$ is the exponential generating function for ordered partitions of a set into k nonempty blocks, and

$$\sum_{k=0}^{\infty} (e^x - 1)^k = \frac{1}{2 - e^x}$$

is the exponential generating function for all ordered partitions of a set.

Now suppose that $f(x)$ is the exponential generating function for a class of labeled objects and that $f(0) = 0$. As we have seen, $f(x)^k$ is the exponential generating function for k-tuples of these objects. Every k-set can be arranged into a k-tuple in $k!$ ways, so $f(x)^k/k!$ is the exponential generating function for k-sets of these objects.

Thus, for example, $(\mathrm{e}^x - 1)^k/k!$ is the exponential generating function for partitions of a set into k blocks. The numbers $S(n,k)$ defined by

$$\frac{(\mathrm{e}^x - 1)^k}{k!} = \sum_{n=0}^{\infty} S(n,k)\frac{x^n}{n!} \tag{9.1}$$

are called *Stirling numbers of the second kind.* If we sum on k we obtain the exponential generating function $\exp(\mathrm{e}^x - 1)$ for all partitions of a set. The coefficients B_n defined by

$$\sum_{n=0}^{\infty} B_n \frac{x^n}{n!} = \mathrm{e}^{\mathrm{e}^x - 1}$$

are called *Bell numbers.*

In general $\mathrm{e}^{f(x)}$ counts sets of labeled objects each counted by $f(x)$. Another important application of this principle (often called the "exponential formula") is to the enumeration of permutations by cycle structure. A permutation may be considered as a set of cycles. If we weight a cycle of length i by u_i and weight a permutation by the product of the weights of its cycles, then the exponential generating function for cycles is

$$\sum_{n=1}^{\infty}(n-1)!\,u_n\frac{x^n}{n!} = \sum_{n=1}^{\infty} u_n\frac{x^n}{n},$$

and thus the exponential generating function for permutations by cycle structure is

$$\exp(\sum_{n=1}^{\infty} u_n x^n/n).$$

If we set $u_n = u$ for all n, then we are counting permutations by the number of cycles, and we obtain the generating function for the (unsigned) Stirling numbers of the first kind,

$$(1-x)^{-u} = \sum_{n=0}^{\infty} u(u+1)\cdots(u+n-1)\frac{x^n}{n!} = \sum_{n=0}^{\infty}\frac{x^n}{n!}\sum_{k=0}^{n} c(n,k)u^k,$$

which we derived in a different way in section 3.

In some cases, there is a simpler expression for $\mathrm{e}^{f(x)}$ than for $f(x)$. For example, any labeled graph is a set of connected labeled graphs. Thus if $g(x)$ is the exponential generating function for connected labeled graphs, then $\mathrm{e}^{g(x)}$ is the exponential generating function for all labeled graphs. But there are $2^{\binom{n}{2}}$ labeled graphs on $[n]$, so

$$g(x) = \log\left(\sum_{n=0}^{\infty} 2^{\binom{n}{2}}\frac{x^n}{n!}\right).$$

Exponential generating functions often satisfy simple differential equations which can be explained combinatorially. If

$$f(x)=\sum_{n=0}^{\infty} f_n \frac{x^n}{n!},$$

then

$$f'(x)=\sum_{n=0}^{\infty} f_{n+1} \frac{x^n}{n!},$$

so an object counted by $f'(x)$ with label set $[n]$ is the same as an object counted by $f(x)$ with label set $[n+1]$. For example, let

$$f(x)=\sum_{n=0}^{\infty} n! \frac{x^n}{n!}=\frac{1}{1-x}$$

be the exponential generating function for permutations (considered as linear arrangements of numbers). Then $f'(x)$ counts permutations of $[n+1]$ in which only the numbers in $[n]$ are considered to be labels. We can consider $n+1$ to be a "marker" that separates the original permutation into a pair of permutations on $[n]$, and we obtain the differential equation $f'(x)=f(x)^2$. This decomposition can be used to obtain more information about permutations, as we shall see next.

A *descent* of the permutation $a_1a_2\cdots a_n$ is an i for which $a_i > a_{i+1}$. It is convenient to count n as a descent also, if $n>0$. Let

$$A(x)=\sum_{n=0}^{\infty} A_n(t) \frac{x^n}{n!}$$

be the exponential generating function for permutations by descents, where a permutation with k descents is weighted t^k. If we take a permutation $\pi=a_1a_2\cdots a_{n+1}$ on $[n+1]$ and remove the element $n+1$, we are left with two permutations, $\pi_1=a_1a_2\cdots a_{j-1}$ and $\pi_2=a_{j+1}\cdots a_{n+1}$, where $a_j=n+1$. The number of descents of π is the sum of the number of descents of π_1 and π_2 unless π_1 is empty, when π has an additional descent. Thus we obtain the differential equation

$$A'(x)=\big(A(x)-1\big)A(x)+tA(x),$$

together with the initial condition $A(0)=1$. The differential equation is easily solved by separation of variables, yielding

$$A(x)=\frac{1-t}{1-te^{(1-t)x}}$$

The polynomials $A_n(t)$ are called *Eulerian polynomials* and their coefficients are called *Eulerian numbers*.

As another example, let us define an *up–down permutation* to be a permutation $a_1a_2\cdots a_n$ satisfying $a_1 < a_2 > a_3 < a_4 \cdots \gtrless a_n$. Let D_n be the number of up–down permutations of $[n]$ and let

$$T(x) = \sum_{n=0}^{\infty} D_{2n+1}\frac{x^{2n+1}}{(2n+1)!}.$$

Removing $2n+1$ from an up–down permutation of $[2n+1]$ for $n \geqslant 1$ leaves a pair of up–down permutations of odd length. Taking into account the exceptional case $n = 0$, we obtain the differential equation $T'(x) = T(x)^2 + 1$, with the initial condition $T(0) = 0$. Solving the differential equation yields $T(x) = \tan x$. The numbers D_{2n+1} are called *tangent numbers.*

For the generating function

$$S(x) = \sum_{n=0}^{\infty} D_{2n}\frac{x^{2n}}{(2n)!},$$

a similar analysis yields the differential equation $S'(x) = T(x)S(x)$, with the initial condition $S(0) = 1$, which has the solution $S(x) = \sec x$. The numbers D_{2n} are called *secant numbers*. We will show that $S(x) = \sec x$ by a different method in section 11.

Another application of exponential generating functions is to the enumeration of labeled rooted trees. Since a rooted tree can be represented as a root together with a set of subtrees, the exponential generating function $t(x)$ for rooted trees satisfies

$$t(x) = xe^{t(x)}.$$

We can solve this equation by the Lagrange inversion formula, and we obtain

$$\frac{t(x)^k}{k!} = \sum_{n=k}^{\infty} n^{n-k}\binom{n-1}{k-1}\frac{x^n}{n!},$$

which for $k = 1$ gives a formula equivalent to Cayley's.

10. Permutations with restricted position

In this and the next several sections we discuss methods for dealing with formulas that involve subtraction. One way to deal with such formulas is to replace them with equivalent formulas having only positive terms. The example we give here is based on the fact that the formula $\sum_k A_k t^k = \sum_k B_k(t-1)^k$ is equivalent to the formula $\sum_k A_k(t+1)^k = \sum_k B_k t^k$.

Theorem 10.1. *Let R be a subset of $[n] \times [n]$. For any permutation π of $[n]$, let $r(\pi)$ be the number of values of $i \in [n]$ for which $(i, \pi(i)) \in R$. Let*

$$a(t) = \sum_{k=0}^{n} a_k t^k = \sum_{\pi\in\mathcal{S}_n} t^{r(\pi)},$$

where $\mathcal{S}_n$ is the set of permutations of $[n]$. Let b_k be the number of k-subsets of R in which no two pairs agree in either coordinate. Then

$$a(t) = \sum_{k=0}^{n} b_k (n-k)!\,(t-1)^k.$$

In particular,

$$a_0 = a(0) = \sum_{k=0}^{n} b_k (n-k)!\,(-1)^k.$$

(Note that b_k is just the number of k-element matchings in the bipartite graph determined by R; see also chapters 3 and 31.)

Proof. We prove that

$$a(t+1) = \sum_{k=0}^{n} b_k (n-k)!\,t^k$$

by counting in two ways pairs (π, Q), in which $\pi \in \mathcal{S}_n$ and $Q \subseteq G(\pi) \cap R$, where $G(\pi) = \{\,(i, \pi(i)) \mid i \in [n]\,\}$. We weight such a pair by $t^{|Q|}$.

First, we have

$$\sum_{(\pi,Q)} t^{|Q|} = \sum_{\pi} \sum_{Q \subseteq G(\pi)\cap R} t^{|Q|} = \sum_{\pi} (t+1)^{|G(\pi)\cap R|} = a(t+1).$$

Second, we have

$$\sum_{(\pi,Q)} t^{|Q|} = \sum_{Q\subseteq R} |\{\,\pi \mid G(\pi) \supseteq Q\,\}|\,t^{|Q|}.$$

If $G(\pi) \supseteq Q$ then Q does not contain two ordered pairs which agree in either coordinate, and if this condition is satisfied, Q can be expanded to the graph of a permutation in $(n - |Q|)!$ ways. Thus the sum is equal to $\sum_{k=0}^{n} b_k (n-k)!\,t^k$. □

Theorem 10.1 is often proved by inclusion–exclusion, which we discuss in section 12. See, e.g., Riordan (1958, chapters 7 and 8).

For our first example, let $R = \{\,(i,i) \mid i \in [n]\,\}$. Then $a(t)$ counts permutations by fixed points. Here $b_k = \binom{n}{k}$, so

$$a(t) = n! \sum_{k=0}^{n} \frac{(t-1)^k}{k!},$$

and in particular, $a_0 = n! \sum_{k=0}^{n} (-1)^k / k!$ is the number of *derangements* (permutations without fixed points) of $[n]$, denoted d_n.

Next we consider the case $R = \{\,(i,j) \mid i - j \equiv 0 \text{ or } 1 \pmod{n}\,\}$, which is the classical *problème des ménages*. Here we can evaluate b_k by a simple trick, but the generalizations in which $i - j \equiv 0, 1, \ldots, s \pmod{n}$ could be solved by the transfer matrix method.

Let us set $p_{2i-1} = (i,i)$ for $1 \leqslant i \leqslant n$, $p_{2i} = (i,i+1)$ for $1 \leqslant i \leqslant n-1$ and $p_{2n} = (n,1)$. Then b_k is the number of k-subsets of $\{p_1, \ldots, p_{2n}\}$ containing no p_i and p_{i+1} (or p_{2n} and p_1). Then as we saw in section 5,

$$b_k = \frac{2n}{2n-k}\binom{2n-k}{k},$$

and thus

$$a_0 = \sum_{k=0}^{n} (-1)^k \frac{2n}{2n-k}\binom{2n-k}{k}(n-k)!.$$

Finally, let us take $R = \{\,(i,j) \mid i > j\,\}$. Then b_k is the Stirling number $S(n, n-k)$. We prove this by giving a bijection between k-subsets of R counted by b_k and partitions of $[n]$ with $n-k$ blocks: to the subset $\{(i_1,j_1),(i_2,j_2),\ldots,(i_k,j_k)\}$ of R counted by b_k we associate the finest partition in which i_s and j_s are in the same block for each s. Thus

$$a(t) = \sum_{k=0}^{n} S(n,n-k)(n-k)!\,(t-1)^k.$$

If we call this polynomial $a_n(t)$, then a straightforward computation using (9.1) shows that

$$\sum_{n=0}^{\infty} a_n(t)\frac{x^n}{n!} = \frac{t-1}{t-e^{(t-1)x.}} = 1 + t^{-1}\left(-1 + \frac{1-t}{1-te^{(1-t)x}}\right) = 1 + \sum_{n=1}^{\infty} t^{-1}A_n(t)\frac{x^n}{n!},$$

where $A_n(t)$ is the Eulerian polynomial.

Similarly, if we had taken $R = \{\,(i,j) \mid i \geqslant j\,\}$, we would have found $a(t) = A_n(t)$ for all n.

Thus for $n \geqslant 1$ the three polynomials

$$\sum_{\pi \in \mathscr{S}_n} t^{1+|\{\,i \mid \pi(i) > \pi(i+1)\,\}|}, \qquad \sum_{\pi \in \mathscr{S}_n} t^{|\{\,i \mid i \geqslant \pi(i)\,\}|} \quad \text{and} \quad \sum_{\pi \in \mathscr{S}_n} t^{1+|\{\,i \mid i > \pi(i)\,\}|},$$

are all equal. A combinatorial proof is easily found through Foata's transformation: For example, if

$$\pi = 5\,7\cdot2\cdot1\,6\cdot3\,8\cdot4\cdot$$

(where the dots represent descents) then Foata's transformation takes π to

$$\pi_1 = (5\,7\cdot)(2\cdot)(1\,6\cdot3\,8\cdot4\cdot),$$

in which occurrences of $\pi(i) > \pi(i+1)$ together with the extra descent at the end have been transformed into occurrences of $i \geqslant \pi_1(i)$.

The variant of Foata's transformation with left–right maxima instead of minima transforms π to

$$\pi_2 = (5)(7\cdot2\cdot1\,6\cdot3)(8\cdot4),$$

in which occurrences of $\pi(i) > \pi(i+1)$ have been transformed into occurrences of $i > \pi_2(i)$.

11. Cancellation

In this section we consider a technique for simplifying sums of positive and negative terms by cancellation. We have two sets A^+ and A^-, which we think of as "positive objects" with sign +1 and "negative objects" with sign −1. We want to find a combinatorial interpretation to $|A^+| - |A^-|$. We do this by finding a partial pairing of positive objects with negative objects; then $|A^+| - |A^-|$ will be equal to the contribution from the unpaired objects.

Theorem 11.1. *Let $A = A^+ \cup A^-$ and suppose that there is a subset B of A and an involution ω defined on $A \setminus B$ which is sign reversing: if $\omega(x)$ is defined, then $x \in A^+$ if and only if $\omega(x) \in A^-$. Then $|A^+| - |A^-| = |A^+ \cap B| - |A^- \cap B|$.*

In most (but not all) applications, B is a subset of either A^+ or A^-.

As an example, we give a combinatorial proof of the identity

$$\sum_{k=0}^{n} (-1)^k \binom{n}{k} \binom{r+k}{m} = (-1)^n \binom{r}{m-n}.$$

Let us first consider the special case $m = 0$, which we may write as

$$\sum_{k=0}^{n} (-1)^k \binom{n}{k} = \begin{cases} 1, & n = 0, \\ 0, & n > 0. \end{cases}$$

It is clear that we should take A to be the set of subsets of $[n]$, with A^+ the subsets of even cardinality and A^- the subsets of odd cardinality. For $n > 0$ we want to find a sign-reversing involution on all of A, so that $B = \emptyset$. Clearly the map given by $\omega(K) = K \triangle \{1\}$ has the right properties, where $\triangle$ denotes the symmetric difference.

Now we consider the general case. Let R be an r-element set disjoint from $[n]$. We may take A to be the set of all pairs (K, M), where K is a subset of $[n]$ and M is an m-subset of $R \cup K$. Then the number of such pairs with $|K| = k$ is $\binom{n}{k}\binom{r+k}{m}$. We take the sign of (K, M) to be $(-1)^{|K|}$. It is not immediately obvious what B should be, but we may try to construct a sign-reversing involution on as large a subset of A as possible, and B will be whatever is left over. Given a pair $(K, M) \in A$, let j be the least element of $[n] \setminus M$ if $[n] \setminus M$ is nonempty. Then we set $\omega((K, M)) = (K \triangle \{j\}, M)$. This is clearly a sign-reversing involution. The only pairs (K, M) for which it is not defined are those for which $[n] \subseteq M$. But if $[n] \subseteq M$ then since $M \cap [n] \subseteq K$, we must have $K = [n]$ and M must consist of $[n]$ together with an $(m-n)$-subset of R. There are $\binom{r}{m-n}$ of these and they all have sign $(-1)^n$. Thus the identity is proved.

For our next example, let D_n be the number of up–down permutations of $[n]$, as defined in section 9. We give a completely different proof that

$$\sum_{n=0}^{\infty} D_{2n} \frac{x^{2n}}{(2n)!} = \sec x. \tag{11.2}$$

If we multiply both sides of (11.2) by $\cos x$ and equate coefficients of $x^{2n}/(2n)!$, we see that (11.2) is equivalent to the recurrence

$$\sum_{k=0}^{n}(-1)^{n-k}\binom{2n}{2k}D_{2k}=\begin{cases}1, & \text{if } n=0,\\ 0, & \text{otherwise.}\end{cases} \tag{11.3}$$

The case $n=0$ of (11.3) is clear. To interpret (11.3) for $n>0$, let A be the set of all ordered pairs (α,β) such that for some subset $S\subseteq[2n]$ of even cardinality, α is an up–down permutation of S and β is the increasing permutation of $[2n]\setminus S$. If $|S|$ has cardinality $2k$, then we give (α,β) the sign $(-1)^k$. Thus for $n=8$, a typical element of A is (1427,3568). Now let $\gamma=(a_1a_2\cdots a_{2k},b_1b_2\cdots b_{2n-2k})$ be an element of A. If $a_{2k}>b_1$ or $k=0$, we define $\omega(\gamma)$ to be $(a_1a_2\cdots a_{2k}b_1b_2,b_3\cdots b_{2n-2k})$ and if $a_{2k}<b_1$ or $k=n$ we define $\omega(\gamma)$ to be $(a_1a_2\cdots a_{2k-2},a_{2k-1}a_{2k}b_1b_2b_3\cdots b_{2n-2k})$. It is clear that ω is a sign-reversing involution defined on all of A, and thus (11.3) is proved. The formula

$$\sum_{n=0}^{\infty}D_{2n+1}\frac{x^{2n+1}}{(2n+1)!}=\tan x$$

can be proved similarly.

Following Zeilberger (1985), we now use a sign-reversing involution to prove the "matrix tree theorem", which gives a determinantal formula for the number of spanning arborescences of a digraph, rooted at a given node. Similar proofs have been found by several people, of whom the first seems to be Temperley (1981).

For each i,j, with $1\leqslant i,j\leqslant n$, let w_{ij} be an arbitrary weight. We define the weight of a digraph on $[n]$ to be the product $\prod w_{ij}$ over all arcs (i,j) of the digraph. We shall find a formula for the sum of the weights of all arborescences on $[n]$, rooted at n. (Then given any digraph D on $[n]$, the number of spanning arborescences of D is obtained by setting w_{ij} equal to 1 for arcs (i,j) in D and to 0 for arcs not in D.)

First we observe that a determinant can be interpreted as a sum of signed weights of digraphs. A *permutation digraph* is a digraph in which every vertex has indegree and outdegree 1, or equivalently, in which every weakly connected component is a directed cycle. Any permutation π of a set corresponds to the permutation digraph in which the arcs are $(i,\pi(i))$, and conversely, every permutation digraph is of this form. Now let M be the matrix $(-w_{ij})_{i,j=1,\ldots,n-1}$. Then the determinant of M is equal to the sum over all permutations π of $[n-1]$ of

$$(\operatorname{sgn}\pi)\prod_{i=1}^{n-1}(-w_{i\pi(i)}). \tag{11.4}$$

This product is clearly, up to sign, the weight of the permutation digraph corresponding to π. Now suppose that π has r cycles, of lengths l_1, l_2, ..., l_r. Then $\operatorname{sgn}\pi=\prod_{i=1}^{r}(-1)^{l_i+1}$ and $(-1)^{n-1}=\prod_{i=1}^{r}(-1)^{l_i}$, so (11.4) is $(-1)^r$ times the weight of the permutation digraph corresponding to π.

Now consider the determinant

$$W = \begin{vmatrix} w_{21} + \cdots + w_{n1} & -w_{21} & \cdots & -w_{n-1,1} \\ -w_{12} & w_{12} + w_{32} + \cdots + w_{n2} & \cdots & -w_{n-1,2} \\ \vdots & \vdots & \ddots & \vdots \\ -w_{1,n-1} & -w_{2,n-1} & \cdots & w_{1,n-1} + \cdots + w_{n-2,n-1} + w_{n,n-1} \end{vmatrix}.$$

This is the determinant of M above, with w_{ii} replaced by

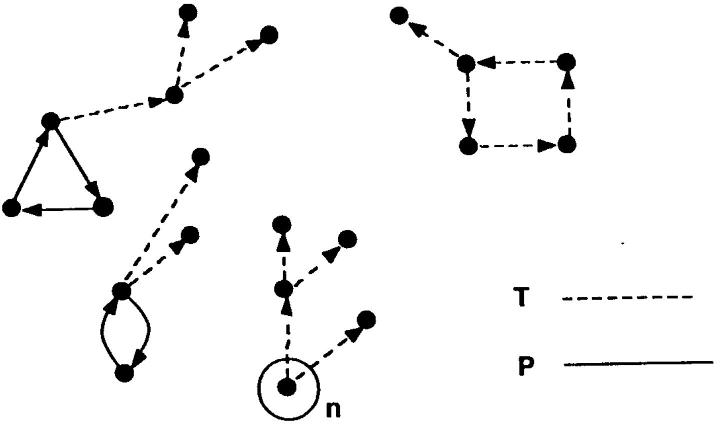
$$-\sum_{\substack{1 \leqslant j \leqslant n \\ j \neq i}} w_{ji}.$$

The digraphs that W counts will be obtained from permutation digraphs by replacing each loop (i,i) with an arc (j,i) for some $j \neq i$ (with $j = n$ allowed), and the sign of such a digraph is $(-1)^r$, where r is the number of cycles of length at least 2 in the original permutation digraph. More precisely, W is the sum of the signed weights of all pairs (P,T) of digraphs on $[n]$ such that:

(1) P is a permutation digraph, with every cycle of length at least 2, on a set of nodes $N_P \subseteq [n-1]$.

(2) T is a digraph without loops on $[n]$ in which every node in $[n-1] \setminus N_P$ has indegree 1 and every node in $N_P \cup \{n\}$ has indegree 0.

The signed weight of the pair (P,T) is $(-1)^r$ times the product of the weights of P and T, where r is the number of cycles of P. Here is a typical pair (P,T):



We now define the sign-reversing involution ω on all pairs (P,T) such that either P or T contains a cycle: take the cycle containing the least vertex and transfer it from P to T or from T to P. Then ω is a weight-preserving sign-reversing involution that cancels all pairs except those in which P is empty and T is an arborescence rooted at n.

Further examples of cancellation can be found in Stanton and White (1986).

12. Inclusion–exclusion

The inclusion–exclusion principle is probably the most well-known technique for dealing with subtraction.

Theorem 12.1. *Let f and g be two functions defined on the subsets of a finite set S such that $f(A)=\sum_{B\subseteq A} g(B)$. Then $g(A)=\sum_{B\subseteq A}(-1)^{|A-B|}f(B)$.*

Proof. We have

$$\sum_{B\subseteq A}(-1)^{|A-B|}f(B)=\sum_{\substack{B\subseteq A\\ C\subseteq B}}(-1)^{|A-B|}g(C)$$

$$=\sum_{C\subseteq A}g(C)\sum_{C\subseteq B\subseteq A}(-1)^{|A-B|}=g(A). \qquad \square$$

A dual form of inclusion–exclusion may be proved the same way as Theorem 12.1:

$$f(A)=\sum_{S\supseteq B\supseteq A}g(B) \text{ if and only if } g(A)=\sum_{S\supseteq B\supseteq A}(-1)^{|B-A|}f(B). \tag{12.2}$$

An important special case of inclusion–exclusion occurs when $f(A)$ and $g(A)$ depend only on $|A|$, so we may write $f(A)=f_{|A|}$ and $g(A)=g_{|A|}$. Then the relation between f and g may be written

$$f_n=\sum_{k=0}^{n}\binom{n}{k}g_k \quad\text{and}\quad g_n=\sum_{k=0}^{n}(-1)^{n-k}\binom{n}{k}f_k.$$

These relations may be expressed in terms of exponential generating functions: if $F(x)=\sum_{n=0}^{\infty}f_nx^n/n!$ and $G(x)=\sum_{n=0}^{\infty}g_nx^n/n!$ then $F(x)=\mathrm{e}^xG(x)$ and $G(x)=\mathrm{e}^{-x}F(x)$.

Another form of inclusion–exclusion is often used: Suppose we have a finite set X of elements, each of which has certain "properties," and let S be the set of all such properties. For each subset A of S let $f(A)$ be the number of elements of X having all the properties in A (and possibly others).

Theorem 12.3. *Let $M_i=\sum_{|A|=i}f(A)$ and let N_i be the number of elements of X having exactly i properties. Then*

$$N_i=\sum_{l\geqslant i}(-1)^{l-i}\binom{l}{i}M_l,$$

and in particular,

$$N_0=M_0-M_1+M_2-\cdots.$$

Proof. For $A \subseteq S$, let $g(A)$ be the number of elements of X having the properties in A and no others. Then $f(A) = \sum_{B \supseteq A} g(B)$, so by inclusion–exclusion, $g(A) = \sum_{B \supseteq A} (-1)^{|B-A|} f(B)$. Thus $N_i = \sum_{|A|=i} g(A)$ and the result follows by a straightforward calculation. □

Our first example of inclusion–exclusion is to permutation enumeration. The *descent set* $D(\pi)$ of a permutation π of $[n]$ is $\{\, i \mid \pi(i) > \pi(i+1) \,\}$. Fix n, and for $A \subseteq [n-1]$, let $g(A)$ be the set of permutations with descent set A. We shall find a simple formula for $f(A) = \sum_{B \subseteq A} g(B)$. Let $A = \{a_1 < a_2 < \cdots < a_k\}$. Then $D(\pi) \subseteq A$ if and only if $\pi(1) < \pi(2) \cdots < \pi(a_1)$, $\pi(a_1+1) < \cdots < \pi(a_2)$, $\ldots$, $\pi(a_k+1) < \cdots < \pi(n)$. To construct such a permutation π, we choose a_1 elements of $[n]$ to be $\{\pi(1), \ldots, \pi(a_1)\}$ and arrange them in increasing order, then choose $a_2 - a_1$ of the remaining elements to be $\{\pi(a_1+1), \ldots, \pi(a_2)\}$, and so on. Thus $f(A)$ is the multinomial coefficient

$$\binom{n}{a_1, a_2 - a_1, \ldots, a_k - a_{k-1}, n - a_k}$$

so $g(A)$ is given explicitly by $g(A) = \sum_{B \subseteq A} (-1)^{|A-B|} f(B)$.

If we set $a_0 = 0$ and $a_{k+1} = n$, then $g(A)$ can be expressed compactly as the determinant

$$n! \left| \frac{1}{(a_j - a_{i-1})!} \right|_{i,j=1,\ldots,k+1}, \tag{12.4}$$

where we interpret $1/r!$ as 0 for $r < 0$. To see this, suppose that (m_{ij}) is an $r \times r$ matrix for which $m_{ij} = 0$ if $j < i - 1$. Then, if $\prod_{i=1}^{r} m_{i\,\pi(i)} \neq 0$, every cycle of π must be of the form $(t\ t-1\ \cdots\ s+1\ s)$. If in addition $m_{i,i-1} = 1$ for $2 \leqslant i \leqslant r$, then the contribution to the determinant $|m_{ij}|$ from the permutation $(t_1\ t_1 - 1 \cdots 2\ 1)(t_2 \cdots t_1 + 1) \cdots (t_l \cdots t_{l-1} + 1)$, where $t_1 < t_2 < \cdots < t_l = r$, is $(-1)^{r-l} m_{1,t_1} m_{t_1+1,t_2} \cdots m_{t_{l-1}+1,r}$. We obtain (12.4) by taking $r = k+1$, $m_{ij} = 1/(a_j - a_{i-1})!$.

As another example, we find a formula for the number c_n of cyclic permutations π of $[n]$ satisfying $\pi(i) \not\equiv i+1 \pmod n$. For any subset A of $[n]$ let $f(A)$ be the number of permutations π with $\pi(i) \equiv i+1 \pmod n$ for all i in A and let $g(A)$ be the number of permutations π with $\pi(i) \equiv i+1 \pmod n$ for all i in A but for no other i. Thus $c_n = g(\emptyset)$. Then it is clear that $f(A) = \sum_{B \supseteq A} g(B)$, so by (12.2), $g(A) = \sum_{B \supseteq A} (-1)^{|B-A|} f(B)$. It is easily seen that $f(A) = (n-1-|A|)!$ for $|A| < n$, with $f([n]) = 1$. Thus

$$c_n = (-1)^n + \sum_{k=0}^{n-1} (-1)^k \binom{n}{k} (n-1-k)!\,.$$

If instead of considering only cyclic permutations, we counted all permutations π satisfying $\pi(i) \not\equiv i+1 \pmod n$, we would have obtained the derangement number d_n. The numbers c_n are closely related to the derangement numbers; it can be shown that $d_n = c_n + c_{n+1}$ and $c_{n+1} = (-1)^{n+1} + \sum_{k=0}^{n} (-1)^{n-k} d_k$.

13. Möbius inversion

Consider the following problem: out of 100 students who are taking Algebra, Biology, and Chemistry, 23 have Algebra and Biology at the same time, 40 have Algebra and Chemistry at the same time, 42 have Biology and Chemistry at the same time, and 15 have all three courses at the same time. How many students have no schedule conflict?

We can solve this problem by inclusion–exclusion. Let U be the set of all 100 students. Let S_1 be the subset of students with an Algebra–Biology conflict, and similarly for S_2 and S_3. Then the answer is

$$|U| - \sum_{i} |S_i| + \sum_{i<j} |S_i \cap S_j| - |S_1 \cap S_2 \cap S_3|.$$

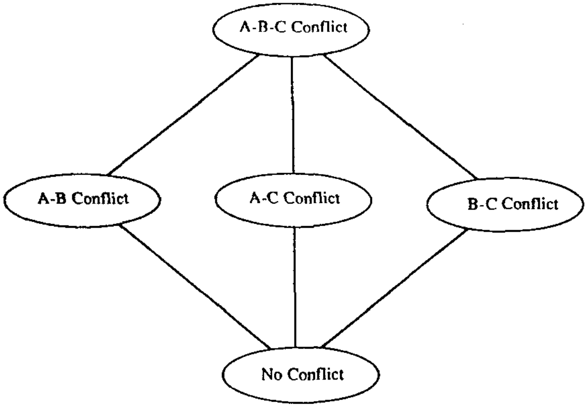
But in this case

$$|S_1 \cap S_2| = |S_1 \cap S_3| = |S_2 \cap S_3| = |S_1 \cap S_2 \cap S_3|,$$

so the formula reduces to

$$|U| - |S_1| - |S_2| - |S_3| + 2|S_1 \cap S_2 \cap S_3| = 25. \tag{13.1}$$

The theory of Möbius inversion explains formulas like (13.1), and in particular explains the significance of the coefficient 2. In this problem there are 5 possibilities for a student's schedule conflict: no conflict, A–B conflict, A–C conflict, B–C conflict, and A–B–C conflict. These conflicts are partially ordered in a natural way as follows:



Then if we let $g(x)$ be the number of students with conflict of type x (but no worse), then we want to determine $g(\text{no conflict})$ given $f(x)$ for all x, where $f(x) = \sum_{y \geqslant x} g(y)$.

In the general situation, we have a finite poset P and two functions f and g on P related by

$$f(x)=\sum_{y\geqslant x} g(y), \tag{13.2}$$

and we want to find the coefficients $m(x,y)$ which express g in terms of f;

$$g(x)=\sum_{y\geqslant x} m(x,y)f(y). \tag{13.3}$$

It is convenient to consider the problem from a slightly different point of view. First let P be a finite poset. The *incidence algebra* $\mathcal{I}(P)$ of P is the set of all complex-valued functions f on $P\times P$ such that $f(x,y)=0$ unless $x\leqslant y$. Addition of these functions is pointwise and multiplication is defined by the formula

$$(fg)(x,y)=\sum_{x\leqslant z\leqslant y} f(x,z)g(z,y).$$

$\mathcal{I}(P)$ is isomorphic to an algebra of matrices in which the rows and columns are indexed by the elements of P; the function f corresponds to the matrix in which the (x,y) entry is $f(x,y)$. If the rows and columns are ordered consistently with P, then these matrices will all be upper triangular. In particular, if $f(x,x)$ is nonzero for all x, then f is invertible.

There are three particularly important elements of the incidence algebra. First there is the identity element δ defined by

$$\delta(x,y)=\begin{cases}1, & \text{if } x=y,\\ 0, & \text{if } x\neq y.\end{cases}$$

Next is the *zeta function* ζ defined by

$$\zeta(x,y)=\begin{cases}1, & \text{if } x\leqslant y,\\ 0, & \text{otherwise}.\end{cases}$$

The *Möbius function* μ of P is the inverse of ζ. By the remark above, μ must exist. An easy way to compute μ is from the recurrence

$$\mu(x,y)=-\sum_{x\leqslant z<y}\mu(x,z),$$

for $x<y$, with the initial condition $\mu(x,x)=1$. This recurrence follows immediately from the formula $\mu\zeta=\delta$.

It is easy to give a formula for $\mu(x,y)$. We have $\zeta^{-1}=(\delta+\zeta-\delta)^{-1}$. It is clear that $(\zeta-\delta)^k(x,y)$ is the number of chains $x=x_0<x_1<\cdots<x_k=y$ and thus is zero for k sufficiently large. So we have the explicit formula

$$\mu=(\delta+(\zeta-\delta))^{-1}=\sum_{k\geqslant 0}(-1)^k(\zeta-\delta)^k,$$

where only finitely many terms on the right are nonzero. If we define the *length* of a chain to be one less than its cardinality, we have P. Hall's theorem:

Theorem 13.4. *$\mu(x,y) = C_0 - C_1 + C_2 - \cdots$, where C_i is the number of chains of length i from x to y.*

Hall's theorem implies that $\mu(x,y)$ depends only on the interval $[x,y] = \{z \mid x \leqslant z \leqslant y\}$. An important, but less obvious, aspect of Hall's theorem is that it provides an interpretation of the Möbius function of a poset P as the reduced Euler characteristic of a topological space associated with P, and thus allows the machinery of algebraic topology to be applied to the study of posets. (See, e.g., Stanley 1986, pp. 120–124 and 137–138.)

Let us return to our original problem. We claim that in (13.3) we should take $m(x,y) = \mu(x,y)$. To see that this works, set

$$\tilde{g}(x) = \sum_{y \geqslant x} \mu(x,y) f(y).$$

Then we have

$$\sum_{y \geqslant x} \tilde{g}(y) = \sum_{y \geqslant x} \sum_{z \geqslant y} \mu(y,z) f(z) = \sum_{z \geqslant x} f(z) \sum_{x \leqslant y \leqslant z} \zeta(x,y)\mu(y,z) = f(x).$$

Since g is uniquely determined by (13.2), we must have $g = \tilde{g}$.

There is a dual form of Möbius inversion in which $y \geqslant x$ is replaced by $y \leqslant x$. We state both forms in the following theorem.

Theorem 13.5. *Let f, g, and h be complex-valued functions on the finite poset P. Then*

(a) $f(x) = \sum_{y \geqslant x} g(y)$ *if and only if* $g(x) = \sum_{y \geqslant x} \mu(x,y) f(y)$;

(b) $h(x) = \sum_{y \leqslant x} g(y)$ *if and only if* $g(x) = \sum_{y \leqslant x} h(y)\mu(y,x)$.

If P and Q are posets, then the product order on $P \times Q$ is given by $(p_1,q_1) \leqslant (p_2,q_2)$ if and only if $p_1 \leqslant p_2$ and $q_1 \leqslant q_2$. The Möbius function of $P \times Q$ is easily expressed in terms of the Möbius functions of P and Q (the straightforward proof is omitted).

Theorem 13.6. *Let P and Q be finite posets. Then*

$$\mu_{P \times Q}\big((p_1,q_1),(p_2,q_2)\big) = \mu_P(p_1,p_2)\mu_Q(q_1,q_2).$$

It is easily seen that if we consider the set $[n]$ as a poset under the usual order, so that it is a chain, then

$$\mu(i,j) = \begin{cases} 1, & \text{if } i = j, \\ -1, & \text{if } i+1 = j, \\ 0, & \text{otherwise.} \end{cases}$$

Since the poset of subsets of a set is a product of 2-element chains, we find that its Möbius function is given by $\mu(A,B) = (-1)^{|B-A|}$ for $A \subseteq B$, which with Theorem 13.5 is the inclusion–exclusion formula.

We now prove two theorems on Möbius functions of lattices. A poset P is a *lattice* if any two elements $x, y \in P$ have a unique join, or least upper bound, denoted $x \vee y$, and a unique meet, or greatest lower bound. We assume that all posets are finite, so any set S of elements of a lattice has a join which we denote by $\bigvee S$. We denote the unique minimal element of a lattice by $\hat{0}$, and the unique maximal element by $\hat{1}$. An *atom* is an element that covers $\hat{0}$.

In our example we computed a Möbius function by using inclusion–exclusion. The next theorem generalizes that example, though we give a different proof.

Theorem 13.7. *Let P be a lattice. Then $\mu(\hat{0}, x) = \sum_S (-1)^{|S|}$, where S ranges over all sets of atoms with join x.*

Proof. For each x in P, let $g(x) = \sum_{\vee S = x} (-1)^{|S|}$, where S ranges over sets of atoms. Define $f(x)$ by $f(x) = \sum_{y \leqslant x} g(y) = \sum_{\vee S \leqslant x} (-1)^{|S|}$. Then, if A is the set of atoms less than or equal to x, we have

$$f(x) = \sum_{S \subseteq A} (-1)^{|S|} = \begin{cases} 1, & \text{if } A = \emptyset \\ 0, & \text{if } A \neq \emptyset \end{cases} = \begin{cases} 1, & \text{if } x = \hat{0}, \\ 0, & \text{if } x \neq \hat{0}. \end{cases}$$

Then by Möbius inversion, $g(x) = \sum_{y \leqslant x} f(y) \mu(y, x) = \mu(\hat{0}, x)$. □

Corollary 13.8. *Under the above hypothesis, if x is not a join of atoms, then $\mu(\hat{0}, x) = 0$.*

Next we prove another basic result on Möbius functions of lattices, called Weisner's theorem.

Theorem 13.9. *Let P be a lattice. Fix a and x in P, with $a > \hat{0}$. Then*

$$\sum_{z \vee a = x} \mu(\hat{0}, z) = 0.$$

Proof. For fixed a, let $g(x) = \sum_{z \vee a = x} \mu(\hat{0}, z)$, and set

$$f(x) = \sum_{y \leqslant x} g(y) = \sum_{z \vee a \leqslant x} \mu(\hat{0}, z).$$

We shall show that $f(x) = 0$ for all x, which implies that $g(x) = 0$. If $a \nleqslant x$ then $f(x)$ is clearly 0. If $a \leqslant x$ then $x \geqslant a > \hat{0}$, so $f(x) = \sum_{z \leqslant x} \mu(\hat{0}, z) = 0$. □

Corollary 13.10. *Let P be a lattice. Suppose that*

(i) *P has a rank function ρ with the property that if a is an atom, then for all x in P, $\rho(a \vee x) \leqslant \rho(x) + 1$.*

(ii) *Every element of P is a join of atoms.*

Then $(-1)^{\rho(\hat{1})} \mu(\hat{0}, \hat{1}) > 0$.

Proof. The assertion is trivially true if $\hat{0} = \hat{1}$. Otherwise, in Theorem 13.9 let a be an atom and take $x = \hat{1}$. Then if $z \vee a = \hat{1}$, z must be $\hat{1}$ or a coatom (of rank $\rho(\hat{1}) - 1$). So $\mu(\hat{0}, \hat{1}) = -\sum_z \mu(\hat{0}, z)$, where the sum is over all coatoms z with $z \vee a = \hat{1}$. The assertion will follow by induction if we can show that a may be chosen so that there is at least one such coatom. But if the sum is empty for all a, then every atom is less than or equal to every coatom, contradicting (ii). □

Lattices satisfying the conditions of Corollary 13.10 are called *geometric lattices*. (There are many other equivalent characterizations of geometric lattices.)

We can use Theorem 13.9 to compute the Möbius function for the lattice L_n of subspaces of the vector space V_n of dimension n over a finite field of q elements. Since the interval $[x, y]$ is isomorphic to L_m, where $m = \dim y - \dim x$, it is sufficient to compute $\mu(\hat{0}, \hat{1})$ in L_n, which we denote by μ_n.

As in Corollary 13.10, let us take a to be an atom and take $x = \hat{1}$. Then, if z is a coatom for which $z \vee a = \hat{1}$, z must be a subspace of V_n of dimension $n-1$ which does not contain a, and the number of these is $\left[{n \atop n-1}\right] - \left[{n-1 \atop n-2}\right] = q^{n-1}$. Thus we have the recurrence $\mu_n = -q^{n-1}\mu_{n-1}$. From this recurrence and the initial condition $\mu_0 = 1$, we obtain

$$\mu_n = (-1)^n q^{\binom{n}{2}}. \tag{13.11}$$

As an application of (13.11), we compute the number $g(x)$ of m-tuples of elements of V_n which span a given subspace x. Let $f(x) = \sum_{y \leqslant x} g(y)$. Then, if $\dim x = d$, we have $f(x) = q^{dm}$, so by Möbius inversion we have

$$g(x) = \sum_{y \leqslant x} f(y)\mu(y, x) = \sum_{k=0}^{d} q^{mk}(-1)^{d-k} q^{\binom{d-k}{2}} \begin{bmatrix} d \\ k \end{bmatrix}.$$

Using (8.5), we can simplify this to

$$g(x) = \prod_{k=0}^{d-1} (q^m - q^k),$$

which can also be found directly. Similarly, the number of m-subsets of V_n with span x (of dimension d) is

$$\sum_{k=0}^{d} \binom{q^m}{k} (-1)^{d-k} q^{\binom{d-k}{2}} \begin{bmatrix} d \\ k \end{bmatrix}.$$

Rota (1964) initiated the systematic use of Möbius functions in combinatorics. Further information about them may be found in chapter 3 of Stanley (1986), and in chapter 31 of this Handbook.

14. Symmetric functions

A formal power series in the variables $x_1, x_2, \ldots, x_n$ is called *symmetric* if it is invariant under any permutation of the variables. It is convenient to work with infinitely many variables, allowing sums such as $x_1 + x_2 + \cdots$. These symmetric formal power series are traditionally (but somewhat misleadingly) called *symmetric functions*.

A symmetric function is *homogeneous of degree k* if every monomial in it has total degree k. It is clear that every symmetric function can be expressed as a (possibly infinite) sum of homogeneous symmetric functions. If we take our coefficients to be complex numbers, then the homogenous symmetric functions of degree k form a vector space, denoted Λ^k. There are several important bases for Λ^k, which are indexed by partitions of k. If $\lambda = (\lambda_1, \ldots, \lambda_n)$ is a partition of k (with the parts listed in decreasing order), then the *monomial symmetric function m_λ* is defined to be the sum of all distinct monomials of the form $x_{i_1}^{\alpha_1} \cdots x_{i_n}^{\alpha_n}$ for permutations $(\alpha_1, \ldots, \alpha_n)$ of λ. It is clear that the m_λ, over all partitions λ of k, form a basis for Λ^k.

For each integer $r \geqslant 0$, the rth *elementary symmetric function e_r* is the sum of all products of r distinct variables, so $e_0 = 1$, and for $r > 0$,

$$e_r = \sum_{i_1 < i_2 < \cdots < i_r} x_{i_1} x_{i_2} \cdots x_{i_r}.$$

For any partition $\lambda = (\lambda_1, \lambda_2, \ldots)$ we define $e_\lambda = e_{\lambda_1} e_{\lambda_2} \cdots$. The "fundamental theorem of symmetric functions" implies that the e_λ over all partitions λ of k form a basis for Λ^k, or equivalently, that every element of Λ^k can be expressed uniquely as a polynomial in the e_r.

The rth *complete symmetric function h_r* is the sum of all monomials of degree r, so $h_0 = 1$ and for $r > 0$,

$$h_r = \sum_{i_1 \leqslant i_2 \leqslant \cdots \leqslant i_r} x_{i_1} x_{i_2} \cdots x_{i_r}.$$

The rth *power sum symmetric function* is

$$p_r = \sum_i x_i^r.$$

For any partition $\lambda = (\lambda_1, \lambda_2, \ldots)$, we define $h_\lambda = h_{\lambda_1} h_{\lambda_2} \cdots$ and $p_\lambda = p_{\lambda_1} p_{\lambda_2} \cdots$.

The generating functions

$$\sum_{r=0}^{\infty} h_r t^r = \prod_{i=1}^{\infty} \frac{1}{1 - x_i t} = \exp\left(\sum_{r=1}^{\infty} \frac{p_r}{r} t^r \right)$$

and

$$\sum_{r=0}^{\infty} e_r t^r = \prod_{i=1}^{\infty} (1 + x_i t) = \left(\sum_{r=0}^{\infty} h_r (-t)^r \right)^{-1}$$

are easy to derive. They imply that e_r can be expressed as a polynomial in the h_i and also in the p_i, and thus $\{h_\lambda\}_{\lambda\vdash k}$ and $\{p_\lambda\}_{\lambda\vdash k}$ are both bases for Λ^k. (Here $\lambda \vdash k$ means that λ is a partition of k.)

There is another important basis for Λ^k which is less obvious. If λ is a partition with n parts, we define the *Schur function* (or *S-function*) s_λ by

$$s_\lambda = \det(h_{\lambda_i - i + j})_{1\leqslant i,j\leqslant n}, \tag{14.1}$$

where we take $h_m = 0$ for $m < 0$.

The Schur functions (in a finite number of variables) arise very naturally from irreducible representations of general linear groups. The irreducible polynomial representations of the general linear group GL_n (over the complex numbers) may be indexed in a natural way by partitions with at most n parts. If χ^λ is the character of the representation associated with λ, then for any matrix M in GL_n with eigenvalues $x_1, x_2, \ldots, x_n$, we have $\chi^\lambda(M) = s_\lambda(x_1, \ldots, x_n)$.

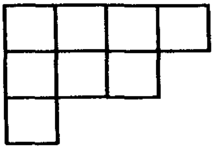
The expansions of the s_λ in the other bases for Λ^k are all interesting. The expansion in elementary symmetric functions is a determinant similar to (14.1).

The expansions of Schur functions in power sum symmetric functions are related to irreducible representations of symmetric groups. There is a natural way of associating to each partition of k an irreducible representation of the symmetric group $\mathcal{S}_k$ (see also chapter 12). Let us denote by χ^λ the character of the representation associated with λ, and by χ^λ_ρ its value at an element of $\mathcal{S}_k$ of cycle type ρ. Then, if λ is a partition of k,

$$s_\lambda = \sum_{\rho\vdash k} \chi^\lambda_\rho \frac{p_\rho}{z_\rho},$$

where $z_\rho = \prod_{i\geqslant 1} i^{m_i} m_i!$ if ρ has m_i parts equal to i.

The coefficients of s_λ (which give its expansion into monomial symmetric functions) have an interesting combinatorial interpretation. The *Ferrers diagram* of a partition λ is an arrangements of cells with λ_i cells, left justified, in the ith row. Thus the Ferrers diagram of the partition $(4, 3, 1)$ is



A *column-strict plane partition of shape* λ is a filling of the Ferrers diagram of λ with positive integers which decrease weakly from left to right and strictly from top to bottom. For example,

5	4	4	1
3	3	1	
1			

is a column-strict plane partition of shape (4,3,1). Then the coefficient of $x_1^{r_1}x_2^{r_2}\cdots x_n^{r_n}$ in s_λ is the number of column-strict plane partitions of shape λ containing r_i entries equal to i.

The *weight* of a plane partition is the sum of its entries. If we set $x_i = q^i$ in s_λ, we get the generating function by weight for column-strict plane partitions of shape λ. There is a very nice explicit formula for this generating function, which can be stated most elegantly in terms of the *hook lengths* of λ. We define the hook length of a cell in a Ferrers diagram to be the number of cells to its right plus the number of cells below it plus one. Thus the hook lengths for the partition (4,3,1) are

6	4	3	1
4	2	1	
1			

Theorem 14.2. *The generating function by weight for column-strict plane partitions of shape λ is*

$$q^{N(\lambda)}\prod_c \frac{1}{1-q^{h(c)}},$$

where the product is over all cells c of the Ferrers diagram of λ, $h(c)$ is the hook length of c, and $N(\lambda) = \sum_i i\lambda_i$.

For the proof of this theorem, and other results on plane partitions, see Stanley (1971) or Macdonald (1979).

One of the most famous theorems of enumerative combinatorics is the theorem of Pólya (1937) on counting orbits under a group action. (See also Pólya and Read 1987.) Pólya's theorem can be stated in several different ways, but one of the most useful is in terms of symmetric functions.

Suppose that a finite group G acts on a finite set A. Then G also acts on functions $f : A \to \mathbb{N}$, where $\mathbb{N}$ is the set of positive integers: for $g \in G$ and $f : A \to \mathbb{N}$, we define $g \cdot f$ by $(g \cdot f)(\alpha) = f(g^{-1}\cdot \alpha)$. We define the *weight* of a function $f : A \to \mathbb{N}$ to be the monomial $\prod_{\alpha\in A} x_{f(\alpha)}$. It is clear that two functions in the same orbit of G have the same weight, so we may define the weight of an orbit to be the weight of any of its elements. Pólya's theorem gives a formula for the sum of the weights of all orbits of functions. We may think of a function $A \to \mathbb{N}$ as a "coloring" of the elements of A, so Pólya's theorem enables us to count colorings which are distinct with respect to the action of G.

Pólya's theorem is a consequence of an elementary result in group theory, often called Burnside's lemma:

Lemma 14.3. *Suppose that a finite group acts on a weighted set X, and that weights are constant on orbits. Define the weight of an orbit to be the weight of any of its* *elements. For each g in G let $\Phi(g)$ be the sum of the weights of the elements of X fixed by G. Then the sum of the weights of the orbits is*

$$\frac{1}{|G|}\sum_{g\in G}\Phi(g).$$

If G acts on a finite set A, then to each element g of G we may associate a permutation π_g of A by $\pi_g(\alpha) = g \cdot \alpha$ for α in A. We define the *cycle index* for the action of G on A to be the symmetric function

$$Z(G) = \frac{1}{|G|}\sum_{g\in G} p_1^{j_1(g)} p_2^{j_2(g)} \cdots, \tag{14.4}$$

where $j_k(g)$ is the number of k-cycles in the cycle decomposition of π_g. We may now state Pólya's theorem:

Theorem 14.5. *The sum of the weights of the orbits of functions on A under the action of G is $Z(G)$.*

Proof. It is not hard to see that a function $f : A \to \mathbb{N}$ is fixed by $g \in G$ if and only if f is constant on each cycle of π_g. Thus the sum of the weights of the functions fixed by g is $p_1^{j_1(g)} p_2^{j_2(g)} \cdots$. Then the theorem follows by applying Lemma 14.3 to the action of G on the set X of functions from A to $\mathbb{N}$. □

One of the simplest applications of Pólya's theorem is to counting equivalence classes of words under the relation of conjugacy introduced in section 5. If we take A to be the set $[n]$, then the functions $A \to \mathbb{N}$ may be identified with words of length n in $\mathbb{N}^*$. Let G be the cyclic group C_n acting in the usual way on $[n]$. Then two words are in the same orbit under the action of C_n if and only if they are conjugates. To evaluate the cycle index of G, let g be a generator for C_n. Then π_{g^m} has d cycles, each of length n/d, where d is the greatest common divisor of m and n. There are $\phi(n/d)$ values of m corresponding to each divisor d of n, where ϕ is Euler's totient function, and thus

$$Z(C_n) = \frac{1}{n}\sum_{d|n}\phi(n/d)p_{n/d}^d. \tag{14.6}$$

In particular, the number of equivalence classes under conjugation of words in $[k]^n$ is obtained by setting $x_1 = x_2 = \cdots = x_k = 1$, $x_i = 0$ for $i > k$, in (14.6), so that $p_i = k$, and (14.6) becomes $n^{-1}\sum_{d|n}\phi(n/d)k^d$.

For a more complicated example, we count isomorphism classes of graphs on n vertices. We start with the action of the symmetric group $\mathscr{S}_n$ on $[n]$. This action yields in a natural way an action on the set A of unordered pairs of distinct elements of $[n]$, which are the edges of the complete graph K_n on $[n]$. Then a function from A to $\mathbb{N}$ may be thought of as a coloring of the edges of K_n. There is a bijection between 2-colorings of edges of K_n and all graphs on $[n]$: given a graph G on $[n]$, we assign an edge of K_n color 1 if it is in G and color 2 if it is not in G. Two graphs are isomorphic if and only if their corresponding 2-colorings of K_n are in the same orbit. Thus to count isomorphism classes of graphs we need only find the cycle index for this action of $\mathcal{S}_n$, then substitute $x_1 = x_2 = 1$; $x_i = 0$ for $i > 2$, which gives $p_i = 2$ for all i.

We shall show that the cycle index is

$$\sum \frac{1}{1^{m_1} m_1! \, 2^{m_2} m_2! \cdots} \prod_k (p_k p_{2k}^{k-1})^{m_{2k}} \cdot \prod_k p_{2k+1}^{k m_{2k+1}} \cdot \prod_k p_k^{k\binom{m_k}{2}} \cdot \prod_{i<j} p_{\mathrm{lcm}(i,j)}^{\gcd(i,j) m_i m_j}, \tag{14.7}$$

where the sum is over all $m_1, m_2, \ldots$ satisfying $m_1 + 2m_2 + \cdots = n$, and lcm and gcd denote the least common multiple and greatest common divisor. To see this, we first observe that the cycle type of π_g for g in $\mathcal{S}_n$ depends only on the cycle type of g. The number of permutations in $\mathcal{S}_n$ with m_i cycles of length i for each i, where $\sum_i i m_i = n$ is

$$\frac{n!}{1^{m_1} m_1! \, 2^{m_2} m_2! \cdots}.$$

For such a permutation g we must determine the cycle type of π_g, the permutation on pairs induced by g.

First we consider pairs in which both elements lie in the same cycle of g. It turns out that we must consider separately cycles of even length and of odd length. In a cycle of g of even length $2k$, the pairs $\{\alpha, g^k(\alpha)\}$ constitute a single cycle of length k; all the other pairs lie in cycles of length $2k$, and there are $k-1$ of these cycles. Thus this cycle of g contributes a factor $p_k p_{2k}^{k-1}$ to the product in (14.7); since there are altogether m_{2k} cycles of this length, their contribution is $(p_k p_{2k}^{k-1})^{m_{2k}}$. For cycles of g of odd length $2k+1$, every pair is in a cycle of π_g of length $2k+1$, and there are k of these, yielding the second product in (14.7).

Next we consider pairs in which the two elements lie in different cycles of g. First suppose that α and β lie in two different cycles of g of the same length k. Then $\{\alpha, \beta\}$ is in a cycle of length k of π_g. The pairs obtained from these two cycles of g will constitute k cycles of π_g, and there are $\binom{m_k}{2}$ ways to choose two cycles of g of length k. This explains the third product in (14.7). Finally, the last product in (14.7) corresponds to the case of a pair of elements from two cycles of g of lengths i and j, with $i < j$. Each pair will lie in a cycle of π_g of length $\mathrm{lcm}(i,j)$. The pairs obtained from these two cycles of g will constitute $\gcd(i,j)$ cycles of π_g, and there are $m_i m_j$ ways to choose two cycles of g of these lengths.

Although (14.7) looks complicated, it is actually useful in computing the number of unlabeled graphs on n vertices, as long as n is not too large. For a comprehensive account of applications of Pólya's theorem to graphical enumeration, see Harary and Palmer (1973).

References

Andrews, G.E.
[1976] *The Theory of Partitions, Encyclopedia of Mathematics*, Vol. 2 (Addison-Wesley, Reading, MA).

Berstel, J., and D. Perrin
[1983] Codes Circulaires, in: *Combinatorics on Words: Progress and Perspectives*, ed. L.J. Cummings (Academic Press Canada, Toronto) pp. 133–166.

Comtet, L.
[1972] Sur les coefficients de l'inverse de la série formelle $\Sigma n!t^n$, *C.R. Acad. Sci. Paris A* **275**, 569–572.
[1974] *Advanced Combinatorics* (Reidel, Dordrecht).

Foata, D.
[1983] Rearrangements of words, in: *Combinatorics on Words, Encyclopedia of Mathematics*, Vol. 17, ed. M. Lothaire (Addison-Wesley, Reading, MA) pp. 184–212.

Garsia, A.M., and S.C. Milne
[1981] A Rogers–Ramanujan bijection, *J. Combin. Theory A* **31**, 289–339.

Goulden, I.P., and D.M. Jackson
[1983] *Combinatorial Enumeration* (Wiley, New York).

Harary, F., and E. Palmer
[1973] *Graphical Enumeration* (Academic Press, New York).

Joyal, A.
[1981] Une théorie combinatoire des séries formelles, *Adv. in Math.* **42**, 1–82.

Labelle, G.
[1981] Une nouvelle démonstration combinatoire des formules d'inversion de Lagrange, *Adv. in Math.* **42**, 217–247.

Macdonald, I.G.
[1979] *Symmetric Functions and Hall Polynomials* (Oxford University Press, Oxford).

McMillan, B.
[1956] Two inequalities implied by unique decipherability, *IRE Trans. Inform. Theory* **IT-2**, 115–116.

Pólya, G.
[1937] Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen, *Acta Math.* **68**, 145–254.

Pólya, G., and R.C. Read
[1987] *Combinatorial Enumeration of Groups, Graphs and Chemical Compounds* (Springer, New York). Contains a translation of Pólya (1937) by Dorothee Aeppli.

Prüfer, H.
[1918] Neuer Beweis eines Satzes über Permutationen, *Arch. Math. Phys.* **27**, 142–144.

Raney, G.
[1960] Functional composition patterns and power series reversion, *Trans. Amer. Math. Soc.* **94**, 441–451.

Riordan, J.
[1958] *An Introduction to Combinatorial Analysis* (Wiley, New York).

Rota, G.-C.
[1964] On the foundations of combinatorial theory I. Theory of Möbius functions, *Z. Wahrscheinlichkeitstheorie u. Verw. Gebiete* **2**, 340–368.

Stanley, R.P.
[1971] Theory and application of plane partitions, parts I and II, *Studia Appl. Math.* **50**, 167–188, 259–279.
[1986] *Enumerative Combinatorics*, Vol. I (Wadsworth, Monterey, CA).

Stanton, D., and D. White
[1986] *Constructive Combinatorics* (Springer, New York).

Temperley, H.N.V.
[1981] *Graph Theory and Applications* (Ellis Horwood/Halsted Press [Wiley], Chichester/New York).

Zeilberger, D.
[1985] A combinatorial approach to matrix algebra, *Discrete Math.* **56**, 61–72.

CHAPTER 22

Asymptotic Enumeration Methods

A.M. ODLYZKO

AT&T Bell Laboratories, 600 Mountain Avenue, Murray Hill, NJ 07974-0636, USA

Contents

HANDBOOK OF COMBINATORICS
Edited by R. Graham, M. Grötschel and L. Lovász

List of examples

Example 3.1. *Dixon's binomial-sum identity.*
Example 5.1. *Sum of primes.*
Example 5.2. *Sums of the partition function.*
Example 5.3. *Sum of* $\exp(-\alpha k^2)$.
Example 5.4. *Bell numbers.*
Example 5.5. *Alternating sum of square roots.*
Example 5.6. *Derangements of n letters.*
Example 5.7. *Balls and cells.*
Example 5.8. *Stirling's formula.*
Example 5.9. *Sum of* $\exp(-\alpha k^2)$.
Example 5.10. *Approximation of Bell numbers.*
Example 6.1. *Rooted labeled trees.*
Example 6.2. *A binomial coefficient sum.*
Example 6.3. *Fibonacci numbers.*
Example 6.4. *Sequences with forbidden subwords.*
Example 6.5. *Polyomino enumeration.*
Example 6.6. *Rooted labeled trees.*
Example 6.7. *Dominant zero for forbidden-subword generating functions.*
Example 6.8. *Waiting times for subwords.*
Example 7.1. *Double set coverings.*
Example 7.2. *Indecomposable permutations.*
Example 8.1. *Lower bound for factorials.*
Example 8.2. *Upper bound for the partition function.*
Example 8.3. *Lower bound for the partition function.*
Example 8.4. *Cycles of permutations.*
Example 8.5. *Permutations with distinct cycle lengths.*
Example 9.1. *Fibonacci numbers.*
Example 9.2. *Sequences with forbidden subblocks.*
Example 9.3. *Two-sided generalized Fibonacci sequences.*
Example 9.4. *An oscillating sequence.*
Example 9.5. *Planted plane trees.*
Example 9.6. *A quadratic recurrence.*
Example 9.7. *Doubly exponential sequences.*
Example 10.1. *Oscillating sequence.*
Example 10.2. *Set partitions with distinct block sizes.*
Example 10.3. *Bernoulli numbers.*
Example 10.4. *Rational function asymptotics.*
Example 10.5. *Permutations without long increasing subsequences.*
Example 10.6. *Sequences with forbidden subblocks.*
Example 10.7. *Coins in a fountain.*
Example 10.8. *Rooted labeled trees.*
Example 11.1. *Partitions with bounded part sizes.*
Example 11.2. *2-regular graphs.*

1. Introduction

Asymptotic enumeration methods provide quantitative information about the rate of growth of functions that count combinatorial objects. Typical questions that these methods answer are: (1) How does the number of partitions of a set of n elements grow with n? (2) How does this number compare to the number of permutations of that set?

There do exist enumeration results that leave nothing to be desired. For example, if a_n denotes the number of subsets of a set with n elements, then we trivially have $a_n = 2^n$. This answer is compact and explicit, and yields information about all aspects of this function. For example, congruence properties of a_n reduce to well-studied number-theory questions. (This is not to say that all such questions have been answered, though!) The formula $a_n = 2^n$ also provides complete quantitative information about a_n. It is easy to compute for any value of n, its behavior is about as simple as possible, and it holds uniformly for all n. However, such examples are extremely rare. Usually, even when there is a formula for the function we are interested in, it is a complicated one, involving summations or recurrences. The purpose of asymptotic methods is to provide simple explicit formulas that describe the behavior of a sequence for large values of indices. There is no satisfactory definition of what is meant by "simple" or by "explicit". However, we can illustrate this concept by some examples. The number of permutations of n letters is given by $b_n = n!$. This is a compact notation, but only in the sense that factorials are so widely used that they have a special symbol. The symbol $n!$ stands for $n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 2 \cdot 1$, and it is the latter formula that has to be used to answer questions about the number of permutations. If one is after arithmetic information, such as the highest power of 7, say, that divides $n!$, one can obtain it from the product formula, but even then some work has to be done. For most quantitative purposes, however, $n! = n \cdot (n-1) \cdot \ldots \cdot 2 \cdot 1$ is inadequate. Since this formula is a product of n terms, most of them large, it is clear that $n!$ grows rapidly, but it is not obvious just how rapidly. Since all but the last term are $\geqslant 2$, we have $n! \geqslant 2^{n-1}$, and since all but the last two terms are $\geqslant 3$, we have $n! \geqslant 3^{n-2}$, and so on. On the other hand, each term is $\leqslant n$, so $n! \leqslant n^n$. Better bounds can clearly be obtained with greater care. The question such estimates raise is: just how far can one go? Can one obtain an estimate for $n!$ that is easy to understand, compute, and manipulate? One answer provided by asymptotic methods is Stirling's formula: $n!$ is asymptotic to $(2\pi n)^{1/2}(n/\mathrm{e})^n$ as $n \to \infty$, which means that the limit as $n \to \infty$ of $n!(2\pi n)^{-1/2}(n/\mathrm{e})^{-n}$ exists and equals 1. This formula is concise and gives a useful representation of the growth rate of $n!$. It shows, for example, that for n large, the number of permutations on n letters is considerably larger than the number of subsets of a set with $\lfloor \frac{1}{2} n \log n \rfloor$ elements.

Another simple example of an asymptotic estimate occurs in the "problème des rencontres" (Comtet 1974). The number d_n of *derangements* of n letters, which is the number of ways of handing back hats to n people so that no person receives his or her own hat, is given by

$$d_n = \sum_{k=0}^{n} (-1)^k \frac{n!}{k!} \,. \tag{1.1}$$

This is a nice formula, yet to compute d_n exactly with it requires substantial effort, since the summands are large, and at first glance it is not obvious how large d_n is. However, we can obtain from (1.1) the asymptotic estimate

$$\frac{d_n}{n!} \to \mathrm{e}^{-1} \quad \text{as } n \to \infty \,. \tag{1.2}$$

To prove (1.2), we factor out $n!$ from the sum in (1.1). We are then left with a sum of rapidly decreasing terms that make up the initial segment of the series

$$\mathrm{e}^{-1} = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \,,$$

and (1.2) follows easily. It can even be shown that d_n is the nearest integer to $\mathrm{e}^{-1}n!$ for all $n \geqslant 1$, see (Comtet 1974). The estimate (1.2) does not allow us to compute d_n, but combined with the estimate for $n!$ cited above it shows that d_n grows like $(2\pi n)^{1/2} n^n \mathrm{e}^{-n-1}$. Further, (1.2) shows that the fraction of all ways of handing out hats that results in every person receiving somebody else's hat is approximately $1/\mathrm{e}$. Results of this type are often exactly what is desired.

Asymptotic estimates usually provide information only about the behavior of a function as the arguments get large. For example, the estimate for $n!$ cited above says only that the ratio of $n!$ to $(2\pi n)^{1/2}(n/\mathrm{e})^n$ tends to 1 as n gets large, and says nothing about the behavior of this ratio for any specific value of n. There are much sharper and more precise bounds for $n!$, and they will be presented in section 3. However, it is generally true that the simpler the estimate, the weaker and less precise it is. There seems to be an unavoidable tradeoff between conciseness and precision. Just about the simplest formula that exactly expresses $n!$ is $n \cdot (n-1) \cdot \ldots \cdot 2 \cdot 1$. (We have to be careful, since there is no generally accepted definition of simplicity, and in many situations it is better to use other exact formulas for $n!$, such as the integral formula $n! = \int_0^\infty t^n \mathrm{e}^{-t}\, \mathrm{d}t$ for the Γ-function. There are also methods for evaluating $n!$ that are somewhat more efficient than the straightforward evaluation of the product.) Any other formula is likely to involve some loss of accuracy as a penalty for simplicity.

Sometimes, the tradeoffs are clear. Let $p(n)$ denote the number of partitions of an integer n. The Rademacher convergent series representation (Andrews 1976, Ayoub 1963) for $p(n)$ is valid for any $n \geqslant 1$:

$$p(n) = \pi^{-1} 2^{-1/2} \sum_{m=1}^{\infty} A_m(n) m^{1/2} \frac{\mathrm{d}}{\mathrm{d}v} (\lambda_v^{-1} \sinh(C m^{-1} \lambda_v)) \Big|_{v=n} \,, \tag{1.3}$$

where

$$C = \pi(2/3)^{1/2}, \quad \lambda_v = (v - 1/24)^{1/2} \,, \tag{1.4}$$

and the $A_m(n)$ satisfy

$$A_1(n) = 1, \quad A_2(n) = (-1)^n \quad \text{for all } n \geqslant 1 \,,$$
$$|A_m(n)| \leqslant m, \quad \text{for all } m, n \geqslant 1 \,,$$

and are easy to compute. Remarkably enough, the series (1.3) does yield the exact integer value of $p(n)$ for every n, and it converges rapidly. (Although this is not directly relevant, we note that using this series to compute $p(n)$ gives an algorithm for calculating $p(n)$ that is close to optimal, since the number of bit operations is not much larger than the number of bits of $p(n)$.) By taking more and more terms, we obtain better and better approximations. The first term in (1.3) shows that

$$p(n) = \pi^{-1}2^{-1/2}\frac{\mathrm{d}}{\mathrm{d}v}(\lambda_v^{-1}\sinh(C\lambda_v))\bigg|_{v=n} + \mathrm{O}(n^{-1}\exp(Cn^{1/2}/2)) \,, \tag{1.5}$$

and if we do not like working with hyperbolic sines, we can derive from (1.5) the simpler (but less precise) estimate

$$p(n) = \frac{1 + \mathrm{O}(n^{-1/2})}{4 \cdot 3^{1/2} n} \, \mathrm{e}^{Cn^{1/2}} \,, \tag{1.6}$$

valid for all $n \geqslant 1$. Unfortunately, exact and rapidly convergent series such as (1.3) occur infrequently in enumeration, and in general we have to be content with poorer approximations.

The advantage of allowing parameters to grow large is that in surprisingly many cases, even when there do exist explicit expressions for the functions we are interested in, this procedure does yield simple asymptotic approximations, when the influence of less important factors falls off. The resulting estimates can then be used to compare numbers of different kinds of objects, decide what are the most common objects in some category, and so on. Even in situations where bounds valid for all parameter values are needed, asymptotic estimates can be used to suggest what form those bounds should take. Usually the error terms in asymptotic estimates can be made explicit (although good bounds often require substantial work), and can be used together with computations of small values to obtain universal estimates. It is common that already for n not much larger than 10 (where n is the basic parameter) the asymptotic estimate is accurate to within a few percentage points, and for $n \geqslant 100$ it is accurate to within a fraction of a percentage point, even though known proofs do not guarantee results as good as this. Therefore the value of asymptotic estimates is much greater than if they just provided a picture of what happens at infinity.

Under some conditions, asymptotic results can be used to prove completely uniform results. For example, if there were any planar maps that were not four-colorable, then almost every large planar map would not be four-colorable, as it would contain one of those small pathological maps. Therefore if it could be proved that most large planar maps are four-colorable, we would obtain a new proof of the four-color theorem that would be more satisfactory to many people than the original one of Haken and Appel. Unfortunately, while this is an attractive idea, no proof of the required asymptotic estimate for the normal chromatic number of planar maps has been found so far.

Asymptotic estimates are often useful in deciding whether an identity is true. If the growth rates of the two functions that are supposed to be equal are different, then the coincidence of initial values must be an accident. There are also more ingenious ways, such as that of Example 13.1, for deducing nonexistence of identities in a wide class from asymptotic information. Sometimes asymptotics is used in a positive way, to suggest what identities might hold.

Simplicity is an important advantage of asymptotic estimates. They are even more useful when no explicit formulas for the function being studied are available, and one has to deal with indirect relations. For example, let T_n be the number of rooted unlabeled trees with n vertices, so that $T_0 = 0$, $T_1 = T_2 = 1$, $T_3 = 2$, $T_4 = 4, \ldots$. No explicit formula for the T_n is known. However, if

$$T(z) = \sum_{n=1}^{\infty} T_n z^n \tag{1.7}$$

is the ordinary generating function of T_n, then Cayley and Pólya showed that

$$T(z) = z \exp\left(\sum_{k=1}^{\infty} T(z^k)/k \right) . \tag{1.8}$$

This functional equation can be derived using the general Pólya–Redfield enumeration method, an approach that is sketched in section 15. Example 15.1 shows how analytic methods can be used to prove, starting with eq. (1.8), that

$$T_n \sim C r^{-n} n^{-3/2} \quad \text{as } n \to \infty , \tag{1.9}$$

where

$$C = 0.4399237\ldots , \quad r = 0.3383219\ldots , \tag{1.10}$$

are constants that can be computed efficiently to high precision. For $n = 20$, $T_n =$ 12, 826, 228, whereas $Cr^{-20}20^{-3/2} = 1.274\ldots \times 10^7$, so asymptotic formula (1.9) is accurate to better than 1%. Thus this approximation is good enough for many applications. It can also be improved easily by adding lower-order terms.

Asymptotic enumeration methods are a subfield of the huge area of general asymptotic analysis. The functions that occur in enumeration tend to be of restricted form (often nonnegative and of regular growth, for example) and therefore the repertoire of tools that are commonly used is much smaller than in general asymptotics. This makes it possible to attempt a concise survey of the most important techniques in asymptotic enumeration. The task is not easy, though, as there has been tremendous growth in recent years in combinatorial enumeration and the closely related field of asymptotic analysis of algorithms, and the sophistication of the tools that are commonly used has been increasing rapidly.

In spite of its importance and growth, asymptotic enumeration has seldom been presented in combinatorial literature at a level other than that of a research paper.

There are several books that treat it (Bender and Williamson 1991, Comtet 1974, Graham et al. 1989, Greene and Knuth 1982, Knuth 1973a,b, 1981, Wilf 1990), but usually only briefly. The only comprehensive survey that is available is the excellent and widely quoted paper of Bender (1974). Unfortunately it is somewhat dated. Furthermore, the last two decades have also witnessed a flowering of asymptotic analysis of algorithms, which was pioneered and popularized by Knuth. Combinatorial enumeration and analysis of algorithms are closely related, in that both deal with counting of particular structures. The methods used in the two fields are almost the same, and there has been extensive cross-fertilization between them. The literature on theoretical computer science, especially on average-case analysis of algorithms, can therefore be used fruitfully in asymptotic enumeration. One notable survey paper in that area is that of Vitter and Flajolet (1990). There are also presentations of relevant methods in the books (Greene and Knuth 1982, Hofri 1987, Knuth 1973a,b, 1981, Kemp 1984). Section 18 is a guide to the literature on these topics.

The aim of this chapter is to survey the most important tools of asymptotic enumeration, point out references for the results and methods that are discussed, and to mention additional relevant papers that have other techniques that might be useful. It is intended for a reader who has already used combinatorial, algebraic, or probabilistic methods to reduce a problem to that of estimating sums, coefficients of a generating function, integrals, or terms in a sequence satisfying some recursion. How such a reduction is to be accomplished will be dealt with sparingly, since it is a large subject that is already covered extensively in other chapters, especially chapter 21. We will usually assume that this task has been done, and will discuss only the derivation of asymptotic estimates.

The emphasis in this chapter is on elementary and analytic approaches to asymptotic problems, relying extensively on explicit generating functions. There are other ways to solve some of the problems we will discuss, and probabilistic methods in particular can often be used instead. We will only make some general remarks and give references to this approach in section 16.

The only methods that will be discussed in detail are fully rigorous ones. There are also methods, mostly from classical applied mathematics (cf. Bender and Orszag 1978) that are powerful and often give estimates when other techniques fail. However, we do not treat them extensively (aside from some remarks in section 16.4) since many of them are not rigorous.

Few proofs are included in this chapter. The stress is on presentation of basic methods, with discussions of their range of applicability, statements of general estimates derivable from them, and examples of their applications. There is some repetitiveness in that several functions, such as $n!$, are estimated several times. The purpose of doing this is to show how different methods compare in their power and ease of use. No attempt is made to present derivations starting from first principles. Some of the examples are given with full details of the asymptotic analysis, to explain the basic methods. Other examples are barely more than statements of results with a brief explanation of the method of proof and a reference to where the proof can be found. The reader might go through this chapter, possibly in a random order, looking for methods that might be applicable to a specific problem, or can look for a category of methods that might fit the problem and start by looking at the corresponding sections.

There are no prerequisites for reading most of this chapter, other than acquaintance with advanced calculus and elementary asymptotic estimates. Many of the results are presented so that they can be used in a cookbook fashion. However, many of the applications require knowledge of complex variables.

Section 2 presents the basic notation used throughout the chapter. It is largely the standard one used in the literature, but it seemed worthwhile summarizing it in one place. Section 3 is devoted to a brief discussion of identities and related topics. While asymptotic methods are useful and powerful, they can often be either augmented or entirely replaced by identities, and this section points out how to use them.

Section 4 summarizes the most important and most useful estimates in combinatorial enumeration, namely those related to factorials and binomial coefficients. Section 5 is the first one to feature an in-depth discussion of methods. It deals with estimates of sums in terms of integrals, summation formulas, and the inclusion–exclusion principle. However, it does not present the most powerful tool for estimation of sums, namely generating functions. These are introduced in section 6, ~~which presents~~ some of the basic properties of, and tools for dealing with, generating functions. While most generating functions that are used in combinatorial enumeration converge at least in some neighborhood of the origin, there are also many nonconvergent ones. Section 7 discusses some estimates that apply to all formal series, but are especially useful for nonconvergent ones.

Section 8 is devoted to estimates for convergent power series that do not use complex variables. While not as powerful as the analytic methods presented later, these techniques are easy to use and suffice in many applications.

Section 9 presents a variety of techniques for determining the asymptotics of recurrence relations. Many of these methods are based on generating functions, and some use analytic methods that are discussed later in the chapter. They are presented at this point ~~because they are~~ basic to combinatorial enumeration, and they also provide an excellent illustration of the power of generating functions.

Section 10 is an introduction to the analytic methods for estimating generating functions. Many of the results mentioned here are common to all introductory complex analysis courses. However, there are also many, especially those in sections 10.4 and 10.5, which are not as well known, and are of special value in asymptotics.

Sections 11 and 12 present the main methods used in estimation of coefficients of analytic functions in a single variable. The basic principle is that the singularities of the generating function that are closest to the origin determine the growth rate of the coefficients. If the function does not grow too fast as it approaches those singularities, the methods of section 11 are usually applicable, while if the growth rate is high, methods of section 12 are more appropriate.

Sections 13–15 discuss extensions of the basic methods of sections 10–12 to multivariate generating functions, integral transforms, and problems that involve a combination of methods.

Section 16 is a collection of miscellaneous methods and results that did not easily fit into any other section, yet are important in asymptotic enumeration. Section 17 discusses the extent to which computer algebra systems can be used to derive asymptotic information. Finally, section 18 is a guide to further reading on asymptotics, since this chapter does not provide complete coverage of the topic.

2. Notation

The symbols O, o, and $\sim$ will have the usual meaning throughout this paper:

$$f(z) = \mathrm{O}(g(z)) \quad \text{as } z \to w \quad \text{means} \quad f(z)/g(z) \quad \text{is bounded as } z \to w\ ;$$

$$f(z) = \mathrm{o}(g(z)) \quad \text{as } z \to w \quad \text{means} \quad f(z)/g(z) \to 0 \quad \text{as } z \to w\ ;$$

$$f(z) \sim g(z) \quad \text{as } z \to w \quad \text{means} \quad f(z)/g(z) \to 1 \quad \text{as } z \to w\ .$$

When an asymptotic relation is stated for an integer variable n instead of z, it will implicitly be taken to apply only for integer values of $n \to w$, and then we will always have $w = \infty$ or $w = -\infty$. An introduction to the use of this notation can be found in Graham et al. (1989). Only a slight acquaintance with it is assumed: enough to see that $(1 + \mathrm{O}(n^{-1/3}))^n = \exp(\mathrm{O}(n^{2/3}))$ and $\log(n + n^{1/2}) = \log(n) + n^{-1/2} - (2n)^{-1} + \mathrm{O}(n^{-3/2})$.

The notation $x \to w^-$ for real w means that x tends to w only through values $x < w$.

Some asymptotic estimates refer to *uniform convergence.* As an example, the statement that $f(z) \sim (1-z)^{-2}$ as $z \to 1$ uniformly in $|\mathrm{Arg}(1-z)| < 2\pi/3$ means that for every $\varepsilon > 0$, there is a $\delta < 0$ such that

$$|f(z)(1-z)^2 - 1| \leqslant \varepsilon$$

for all z with $0 < |1-z| < \delta$, $|\mathrm{Arg}(1-z)| < 2\pi/3$. This is an important concept, since lack of uniform convergence is responsible for many failures of asymptotic methods to yield useful results.

Generating functions will usually be written in the form

$$f(z) = \sum_{n=0}^{\infty} f_n z^n\ , \tag{2.1}$$

and we will use the notation $[z^n]f(z)$ for the coefficient of z^n in $f(z)$, so that if $f(z)$ is defined by (2.1), $[z^n]f(z) = f_n$. For multivariate generating functions, $[x^m y^n]f(x,y)$ will denote the coefficient of $x^m y^n$, and so on. If a_n denotes a sequence whose asymptotic behavior is to be studied, then in combinatorial enumeration one usually uses either the *ordinary generating function* $f(z)$ defined by (2.1) with $f_n = a_n$, or else the *exponential generating function* $f(z)$ defined by (2.1) with $f_n = a_n/n!$. In this chapter we will not be concerned with the question of which type of generating function is best in a given context, but will assume that a generating function is given, and will concentrate on methods of extracting information about the coefficients from the form we have.

Asymptotic series, as defined by Poincaré, are written as

$$f_n \sim \sum_{k=0}^{\infty} a_k n^{-k} , \tag{2.2}$$

and mean that for every $K \geqslant 0$,

$$f_n = \sum_{k=0}^{K} a_k n^{-k} + \mathrm{O}(n^{-K-1}) \quad \text{as } n \to \infty . \tag{2.3}$$

The constant implied by the O-notation may depend on K. It is unfortunate that the same symbol is used to denote an asymptotic series and an asymptotic relation, as defined in the first paragraph of this section. Confusion should be minimal, though, since asymptotic relations will always be written with an explicit statement of the limit of the argument.

The notation $f(z) \approx g(z)$ will be used to indicate that $f(z)$ and $g(z)$ are in some vague sense close together. It is used in this chapter only in cases where a precise statement would be cumbersome and would not help in explaining the essence of the argument.

All logarithms will be natural ones to base e unless specified otherwise, so that $\log 8 = 2.0794\ldots$, $\log_2 8 = 3$. The symbol $\lfloor x \rfloor$ denotes the greatest integer $\leqslant x$. The notation $x \to 1^-$ means that x tends to 1, but only from the left, and similarly, $x \to 0^+$ means that x tends to 0 *only* from the right, through positive values.

3. Identities, indefinite summations, and related approaches

Asymptotic estimates are useful, but often they can be avoided by using other methods. For example, the asymptotic methods presented later yield estimates for $\binom{n}{k}2^k$ as k and n vary, which can be used to estimate accurately the sum of $\binom{n}{k}2^k$ for n fixed and k running over the full range from 0 to n. That is a general and effective process, but somewhat cumbersome. On the other hand, by the binomial theorem,

$$\sum_{k=0}^{n} \binom{n}{k} 2^k = (1+2)^n = 3^n . \tag{3.1}$$

This is much more satisfactory and simpler to derive than what could be obtained from applying asymptotic methods to estimate individual terms in the sum. However, such identities are seldom available. There is nothing similar that can be applied to

$$\sum_{k \leqslant n/5} \binom{n}{k} 2^k , \tag{3.2}$$

and we are forced to use asymptotic methods to estimate this sum.

Recognizing when some combinatorial identity might apply is not easy. The literature on this subject is huge, and some of the references for it are Gould (1972), Gradshteyn and Ryzhik (1965), Hansen (1975), Jolley (1961), and Riordan (1968). Many of the books listed in the references are useful for this purpose. Generating functions (see section 6) are one of the most common and powerful tools for proving identities. Here we only mention two recent developments that are of significance for both theoretical and practical reasons. One is Gosper's algorithm for indefinite hypergeometric summation (Gosper 1978, Graham et al. 1989). Given a sequence $a_1, a_2, \ldots$, Gosper's algorithm determines whether the sequence of partial sums

$$b_n = \sum_{k=1}^{n} a_k\,, \quad n = 1, 2, \ldots \tag{3.3}$$

has the property that b_n/b_{n-1} is a rational function of n, and if it is, it gives an explicit form for b_n. We note that if b_n/b_{n-1} is a rational function of n, then so is

$$\frac{a_n}{a_{n-1}} = \frac{b_n/b_{n-1} - 1}{1 - b_{n-2}/b_{n-1}}\,. \tag{3.4}$$

Therefore Gosper's algorithm should be applied only when a_n/a_{n-1} is rational.

The other recent development is the Wilf–Zeilberger method for proving combinatorial identities (Wilf and Zeilberger 1990, 1992). Given a conjectured identity, it provides an algorithmic procedure for verifying it. This method succeeds in a surprisingly wide range of cases. Typically, to prove an identity of the form

$$\sum_k U(n,k) = S(n)\,, \quad n \geqslant 0\,, \tag{3.5}$$

where $S(n) \neq 0$, Wilf and Zeilberger define $F(n,k) = U(n,k)/S(n)$ and search for a rational function $R(n,k)$ such that if $G(n,k) = R(n,k)F(n,k-1)$, then

$$F(n+1,k) - F(n,k) = G(n,k+1) - G(n,k) \tag{3.6}$$

holds for all integers n, k with $n \geqslant 0$, and such that

(i) for each integer k, the limit

$$f_k = \lim_{n\to\infty} F(n,k) \tag{3.7}$$

exists and is finite;

(ii) for each integer $n \geqslant 0$, $\lim_{k\to+\infty} G(n,k) = 0$;

(iii) $\lim_{k\to-\infty} \sum_{n=0}^{\infty} G(n,k) = 0$.

If all these conditions are satisfied, and eq. (3.5) holds for $n = 0$, then it holds for all $n \geqslant 0$.

Example 3.1 (*Dixon's binomial-sum identity*). This identity states that

$$\sum_k (-1)^k \binom{n+b}{n+k}\binom{b+c}{b+k}\binom{n+c}{c+k} = \frac{(n+b+c)!}{n!\,b!\,c!}\,. \tag{3.8}$$

This can be proved by the Wilf–Zeilberger method by taking

$$R(n,k) = \frac{(b+1-k)(c+1-k)}{2(n+k)(n+b+c+1)} \tag{3.9}$$

and verifying that the conditions above hold. ⊠

The Wilf–Zeilberger method requires finding a rational function $R(n,k)$ that satisfies the properties listed above. This is often hard to do, especially by hand. Gosper's algorithm leads to a systematic procedure for constructing such $R(n,k)$.

To conclude this section, we mention that useful resources when investigating sequences arising in combinatorial settings are the books of Sloane (1973) and Sloane and Plouffe (1995), which lists several thousand sequences and gives references for them. Section 17 mentions some software systems that are useful in asymptotics.

4. Basic estimates: factorials and binomial coefficients

No functions in combinatorial enumeration are as ubiquitous and important as the factorials and the binomial coefficients. In this section we state some estimates for these quantities, which will be used throughout this chapter and are of widespread applicability. Several different proofs of some of these estimates will be sketched later.

The basic estimate, from which many others follow, is that for the factorial. As was mentioned in the introduction, the basic form of Stirling's formula is

$$n! \sim (2\pi n)^{1/2} n^n \mathrm{e}^{-n} \quad \text{as } n \to \infty\,. \tag{4.1}$$

This is sufficient for many enumeration problems. However, when necessary one can draw on much more accurate estimates. For example eq. 6.1.38 in NBS (1970) gives

$$n! = (2\pi n)^{1/2} n^n \exp(-n + \theta/(12n)) \tag{4.2}$$

for all $n \geqslant 1$, where $\theta = \theta(n)$ satisfies $0 < \theta < 1$. More generally, there is Stirling's asymptotic expansion:

$$\log\{n!(2\pi n)^{-1/2} n^{-n} \mathrm{e}^{n}\} \sim \frac{1}{12n} - \frac{1}{360n^3} + \cdots\,. \tag{4.3}$$

[This is an asymptotic series in the sense of eq. (2.2), and there is no convergent expansion for $\log\{n!(2\pi n)^{-1/2} n^{-n} \mathrm{e}^{n}\}$ as a power series in n^{-1}.] Further terms in the expansion (4.3) can be obtained, and they involve Bernoulli numbers. In most references, such as eq. 6.1.37 or 6.1.40 of NBS (1970), Stirling's formula is presented for $\Gamma(x)$, where Γ is Euler's gamma function. Expansions for $\Gamma(x)$ translate readily into ones for $n!$ because $n! = \Gamma(n+1)$.

Stirling's approximation yields the expansion

$$\binom{2n}{n} = \frac{4^n}{(\pi n)^{1/2}} \left\{1 - \frac{1}{8n} + \frac{1}{128n^2} + \frac{5}{1024n^3} + \mathrm{O}(n^{-4})\right\}\,. \tag{4.4}$$

A less precise but still useful estimate is

$$\binom{n}{\lfloor n/2 \rfloor} \sim \left(\frac{2}{\pi n}\right)^{1/2} 2^n \quad \text{as } n \to \infty . \tag{4.5}$$

This estimate is used frequently. The binomial coefficients are *symmetric*, so that $\binom{n}{k} = \binom{n}{n-k}$ and *unimodal*, so that for a fixed n and k varying, the $\binom{n}{k}$ increase monotonically up to a peak at $k = \lfloor n/2 \rfloor$ (which is unique for n even and has two equal high points at $k = (n \pm 1)/2$ for n odd) and then decrease.

More important than eq. (4.5) are expansions for general binomial coefficients. Equation (4.2) shows that for $1 \leqslant k \leqslant n-1$,

$$\begin{aligned}\binom{n}{k} &= \frac{n!}{k!(n-k)!} \\ &= \left\{\frac{n}{2\pi k(n-k)}\right\}^{1/2} \frac{n^n}{k^k (n-k)^{n-k}} \exp\left(O\left(\frac{1}{k} + \frac{1}{n-k}\right)\right) \\ &= \left\{\frac{n}{2\pi k(n-k)}\right\}^{1/2} \exp\left(nH\left(\frac{k}{n}\right) + O\left(\frac{1}{k} + \frac{1}{n-k}\right)\right) ,\end{aligned} \tag{4.6}$$

where

$$H(x) = -x \log x - (1-x)\log(1-x) \tag{4.7}$$

is the entropy function. (We set $H(0) = H(1) = 0$ to make $H(x)$ continuous for $0 \leqslant x \leqslant 1$.) Simplifying further, we obtain

$$\binom{n}{k} = \exp(nH(k/n) + O(\log n)) , \tag{4.8}$$

an estimate that is valid for all $0 \leqslant k \leqslant n$. In many situations it suffices to use the weaker but simpler bound

$$\binom{n}{k} \leqslant \left(\frac{n\mathrm{e}}{k}\right)^k , \quad 0 \leqslant k \leqslant n . \tag{4.9}$$

Approximations of this form are used frequently in information theory and other fields.

A general estimate that can be derived by totally elementary methods, without recourse to Stirling's formula, is

$$\binom{n}{k}\binom{n}{\lfloor n/2 \rfloor}^{-1} = \exp(-2(k-n/2)^2/n + O(|k-n/2|^3/n^2)) , \tag{4.10}$$

valid for $|k - n/2| \leqslant n/4$, say. It is most useful for $|k - n/2| = o(n^{2/3})$, since the error term is small then. Similarly,

$$\binom{n}{k+r} \sim \binom{n}{k}\left(\frac{n-k}{k}\right)^r \quad \text{as } n \to \infty , \tag{4.11}$$

uniformly in k provided r (which may be negative) satisfies $r^2 = o(k)$ and $r^2 = o(n-k)$. Further, we have

$$(n+k)! \sim n^k \exp(k^2/(2n))n! \quad \text{as } n \to \infty , \tag{4.12}$$

again uniformly in k provided $k = o(n^{2/3})$.

5. Estimates of sums and other basic techniques

When encountering a combinatorial sum, the first reaction should always be to check whether it can be simplified by use of some identity. If no identity for the sum is found, the next step should be to try to transform the problem to eliminate the sum. Usually we are interested not in single isolated sums, but parametrized families of them, such as

$$b_n = \sum_k a_n(k) , \tag{5.1}$$

and it is the asymptotic behavior of the b_n as $n \to \infty$ that is desired. A standard and well-known technique (named the "snake-oil" method by Wilf 1990) for handling such cases is to form a generating function $f(z)$ for the b_n, use the properties of the $a_n(k)$ to obtain a simple form for $f(z)$, and then obtain the asymptotics of the b_n from the properties of $f(z)$. This method will be presented briefly in section 6. In this section we discuss what to do if those two approaches fail. Sometimes the methods to be discussed can also be used in a preliminary phase to obtain a rough estimate for the sum. This estimate can then be used to decide which identities might be true, or what generating functions to form.

There are general methods for dealing with sums (cf. Knopp 1971), many of which are used in asymptotic enumeration. A basic technique of this type is summation by parts. Often sums to be evaluated can be expressed as

$$\sum_{j=1}^{n} a_j b_j \quad \text{or} \quad \sum_{j=1}^{\infty} a_j b_j ,$$

where the b_j, say, are known explicitly or behave smoothly, while the a_j by themselves might not be known well, but the asymptotics of

$$A(k) = \sum_{j=1}^{k} a_j \tag{5.2}$$

are known. Summation by parts relies on the identity

$$\sum_{j=1}^{n} a_j b_j = \sum_{k=1}^{n-1} A(k)(b_k - b_{k+1}) + A(n)b_n . \tag{5.3}$$

Example 5.1 (*Sum of primes*). Let

$$S_n = \sum_{p \leqslant n} p\ , \tag{5.4}$$

where p runs over the primes $\leqslant n$. The Prime Number Theorem (Ayoub 1963) states that the function

$$\pi(x) = \sum_{p \leqslant x} 1 \tag{5.5}$$

satisfies

$$\pi(x) \sim \frac{x}{\log x} \quad \text{as } x \to \infty\ . \tag{5.6}$$

(More precise estimates are available, but we will not use them.) We rewrite

$$S_n = \sum_{j=1}^{n} a_j b_j\ , \tag{5.7}$$

where

$$a_j = \begin{cases} 1 & j \text{ is prime}\ , \\ 0 & \text{otherwise}\ , \end{cases} \tag{5.8}$$

and $b_j = j$ for all j. Then $A(k) = \pi(k)$ and summation by parts yields

$$S_n = \sum_{k=1}^{n-1} -\pi(k) + \pi(n)n\ . \tag{5.9}$$

Since

$$\sum_{k=1}^{n-1} \pi(k) \sim \sum_{k=2}^{n-1} \frac{k}{\log k} \sim \frac{n^2}{2\log n} \quad \text{as } n \to \infty\ , \tag{5.10}$$

we have

$$S_n \sim \frac{n^2}{2\log n} \quad \text{as } n \to \infty\ . \tag{5.11}$$

⊠

Summation by parts is used most commonly in situations like those of Example 5.1, to obtain an estimate for one sum from that of another.

Summation by parts is often easiest to carry out, both conceptually and notationally, by using integrals. If we let

$$A(x) = \sum_{k \leqslant x} a_k\ , \tag{5.12}$$

then $A(x) = A(n)$ for $n \leqslant x < n+1$. Suppose that $b_k = b(k)$ for some continuously differentiable function $b(x)$. Then

$$b_k - b_{k+1} = -\int_k^{k+1} b'(x)\,\mathrm{d}x \ , \tag{5.13}$$

and we can rewrite eq. (5.3) as

$$\sum_{j=1}^{n} a_j b_j = A(n)b(n) - \int_1^n A(x)b'(x)\,\mathrm{d}x \ . \tag{5.14}$$

[One can apply similar formulas even when the b_j are not smooth, but this usually requires Riemann–Stieltjes integrals, cf. Apostol (1957).] The approximation of sums by integrals that appears in (5.14) is common, and will be treated at length later.

5.1. Sums of positive terms

Sums of positive terms are extremely common. They can usually be handled with only a few basic tools. We devote substantial space to this topic because it is important and because the simplicity of the methods helps in illustrating some of the basic principles of asymptotic estimation, such as approximation by integrals, neglecting unimportant terms, and uniform convergence. For readers not familiar with asymptotic methods, working through the examples of this section is a good exercise that will make it easier to learn other techniques later.

Typical sums are of the form

$$b_n = \sum_k a_n(k) \ , \qquad a_n(k) \geqslant 0 \ , \tag{5.15}$$

where k runs over some range of summation, often $0 \leqslant k \leqslant n$ or $0 \leqslant k < \infty$, and the $a_n(k)$ may be given either explicitly or only through an asymptotic approximation. What is desired is the asymptotic behavior of b_n as $n \to \infty$. Usually the $a_n(k)$ for n fixed are unimodal, so that either (i) $a_n(k) \leqslant a_n(k+1)$ for all k in the range, or (ii) $a_n(k) \geqslant a_n(k+1)$ for all k, or (iii) $a_n(k) \leqslant a_n(k+1)$ for $k \leqslant k_0$, and $a_n(k) \geqslant a_n(k+1)$ for $k > k_0$. The single most important task in estimating b_n is usually to find the maximal $a_n(k)$. This can be done either by combinatorial means (involving knowledge of where the $a_n(k)$ come from), by asymptotic estimation of the $a_n(k)$, or (most common when the $a_n(k)$ are expressed in terms of factorials or binomial coefficients) by finding where the ratio $a_n(k+1)/a_n(k)$ is close to 1. If $a_n(k+1)/a_n(k) < 1$ for all k, then we are in case (ii) above, and if $a_n(k+1)/a_n(k) > 1$ for all k, we are in case (i). If there is a k_0 in the range of summation such that $a_n(k_0+1)$ is close to $a_n(k_0)$, then we are almost certainly in case (iii) and the peak occurs at some k close to k_0. The different cases are illustrated in the examples presented later in this section.

Once $\max a_n(k) = a_n(k_0)$ has been found, the next task is to show that most of the terms in the sum are insignificant. For example, if the sum in eq. (5.15) is over $0 \leqslant k \leqslant n$, and if $a_n(0) = 1$ is the largest term, then

$$\sum_{\substack{k=0 \\ a_n(k)<n^{-2}}}^{n} a_n(k) < n^{-1} ,$$

which is negligible if we are only after a rough approximation to b_n, say of the form $b_n \sim c_n$ as $n \to \infty$, or even $b_n = c_n(1 + \mathrm{O}(n^{-1}))$ as $n \to \infty$. Once the small terms have been discarded, we are usually left with a short range of summation. It can happen that this range is extremely short, and the maximal term $a_n(k_0)$ is much larger than any of its neighbors to the extent that $b_n \sim a_n(k_0)$ as $n \to \infty$. More commonly, the number of terms that contribute significantly to b_n does grow as $n \to \infty$, but slowly. Their contribution, relative to that of the maximal term $a_n(k_0)$, can usually be estimated by some simple function of $k - k_0$, and the sum of all of them approximated by an explicit integral. This method is sometimes referred to as Laplace's method for sums (in analogy to Laplace's method for estimating integrals, mentioned in section 5.5, which proceeds in a similar spirit). There is extensive discussion of this method in de Bruijn (1958).

Example 5.2 (*Sums of the partition function*). We estimate

$$U_n = \sum_{k=1}^{n} p(k)^k , \tag{5.16}$$

where $p(k)$ is the number of partitions of k. Since any partition of $m - 1$, say one with c_j parts of size j, can be transformed into a partition of m with $c_1 + 1$ parts of size 1, and c_j of size j for $j \geqslant 2$, we have $p(m) \geqslant p(m-1)$ for all $m \geqslant 2$. Therefore the largest term in the sum in (5.16) is the one with $k = n$. If the only estimate for $p(k)$ that we have is the one given by (1.6), then

$$p(n)^n = \exp(Cn^{3/2} - n\log(4 \cdot 3^{1/2}n) + \mathrm{O}(n^{1/2})) . \tag{5.17}$$

Since the constant implied by the O-symbol is not specified, this estimate is potentially larger than $p(n)^n$ by a factor of $\exp(cn^{1/2})$, so we can only obtain asymptotics of $\log p(n)^n$, not of $p(n)^n$ itself. This also means that rough estimates of U_n follow easily from (5.17). Since $p(k)^k \leqslant p(n)^n$ for all $k < n$, and there are n terms in the sum, we have $p(n)^n \leqslant U_n \leqslant np(n)^n$, and because of the large error term in (5.17), we obtain

$$U_n = \exp(Cn^{3/2} - n\log(4 \cdot 3^{1/2}n) + \mathrm{O}(n^{1/2})) . \tag{5.18}$$

Thus the use of the poor estimate (1.6) for $p(n)$ means that we can obtain only a crude estimate for U_n, and there is no need for careful analysis.

Instead of (1.6) we can use the more refined estimate (1.5). Let q_n denote the first term on the right side of (1.5). Then we have

$$p(n) = q_n + \mathrm{O}(n^{-1}\exp(Cn^{1/2}/2)) = q_n(1 + \mathrm{O}(\exp(-Cn^{1/2}/2))) , \tag{5.19}$$

so

$$p(n)^n = q_n^n(1+\mathrm{O}(n\exp(-Cn^{1/2}/2))) = q_n^n(1+\mathrm{O}(\exp(-Cn^{1/2}/3))) , \tag{5.20}$$

say. Also, for some $\varepsilon > 0$ we find from eq. (1.5) [or eq. (1.6)] that for large n

$$q_{n-1} < q_n - \varepsilon n^{-1/2} q_n .$$

Thus for large n,

$$\begin{aligned} q_{n-1}^{n-1} &< q_n^{n-1}(1-\varepsilon n^{-1/2})^{n-1} \\ &< q_n^n \exp(-\varepsilon n^{1/2}/2) , \end{aligned}$$

and therefore

$$\sum_{k=1}^{n-1} p(k)^k \leqslant (n-1)p(n-1)^{n-1} < q_n^n \exp(-\varepsilon n^{1/2}/3) .$$

Thus we obtain

$$U_n = q_n^n(1+\mathrm{O}(\exp(-\delta n^{1/2}))) \tag{5.21}$$

for some $\delta > 0$.

The estimates of U_n presented above relied on the observation that the last term in the sum (5.16) defining U_n is much larger than the sum of all the other terms. This does not happen often. A more typical example is presented by

$$T_n = \sum_{k=1}^{n} p(k) . \tag{5.22}$$

As was noted before, $p(n)$ is larger than any of the other terms, but not by enough to dominate the sum. We therefore try the other approaches that were listed at the beginning of this section. We use only the estimate (1.6). Since $(1-x)^{1/2} < 1 - x/2$ for $0 \leqslant x \leqslant 1$, we find that for large n,

$$\begin{aligned} \textstyle\sum_{k<n-n^{2/3}} p(k) &\leqslant np(n-\lceil n^{2/3}\rceil) \\ &\leqslant \exp(C(n-\lceil n^{2/3}\rceil)^{1/2}) \\ &\leqslant \exp(Cn^{1/2} - Cn^{1/6}/2) \\ &= \mathrm{O}(p(n)\exp(-Cn^{1/6}/3)) . \end{aligned} \tag{5.23}$$

Thus most of the values of k contribute a negligible amount to the sum. For $k = n - j$, $0 \leqslant j \leqslant n^{2/3}$, we find that

$$p(n-j)/p(n) = (1+\mathrm{O}(n^{-1/3}))\exp(C(n-j)^{1/2} - Cn^{1/2}) .$$

Since

$$\begin{aligned} (n-j)^{1/2} &= n^{1/2} - jn^{-1/2}/2 + \mathrm{O}(j^2 n^{-3/2}) , \\ p(n-j)/p(n) &= \exp(-Cjn^{-1/2}/2 + \mathrm{O}(n^{-1/6})) \\ &= (1+\mathrm{O}(n^{-1/6}))\exp(-Cjn^{-1/2}/2) . \end{aligned} \tag{5.24}$$

Thus the ratios $p(n-j)/p(n)$ decrease geometrically, and so

$$p(n)^{-1} \sum_{0 \leqslant j \leqslant n^{2/3}} p(n-j) = \frac{(1+\mathrm{O}(n^{-1/6}))}{1-\exp(-Cn^{-1/2}/2)} = 2C^{-1}n^{1/2}(1+\mathrm{O}(n^{-1/6})) . \tag{5.25}$$

Therefore, combining all the estimates,

$$T_n = \sum_{k=1}^{n} p(k) = \frac{1+\mathrm{O}(n^{-1/6})}{2 \cdot C \cdot 3^{1/2} \cdot n^{1/2}} \mathrm{e}^{Cn^{1/2}} . \tag{5.26}$$

The $\mathrm{O}(n^{-1/6})$ error term above can easily be improved with a little more care to $\mathrm{O}(n^{-1/2})$, even if we continue to rely only on (1.6). ⊠

Before presenting further examples, we discuss some of the problems that can arise even in the simple setting of estimating positive sums. We then introduce the basic technique of approximating sums by integrals.

The lack of uniform convergence is a frequent cause of incorrect estimates. If $a_n(k) \sim c_n(k)$ for each k as $n \to \infty$, it does not necessarily follow that

$$b_n = \sum_k a_n(k) \sim \sum_k c_n(k) \quad \text{as } n \to \infty . \tag{5.27}$$

A simple counterexample is given by $a_n(k) = \binom{n}{k}$ and $c_n(k) = \binom{n}{k}(1+k/n)$. To conclude that (5.27) holds, it is usually necessary to know that $a_n(k) \sim c_n(k)$ as $n \to \infty$ uniformly in k. Such uniform convergence does hold if we replace $c_n(k)$ in the counterexample above by $c'_n(k) = \binom{n}{k}(1+k/n^2)$, for example.

There is a general principle that sums of terms that vary smoothly with the index of summation should be replaced by integrals, so that for $\alpha > 0$, say,

$$\sum_{k=1}^{n} k^{\alpha} \sim \int_1^{n+1} u^{\alpha}\, \mathrm{d}u \quad \text{as } n \to \infty . \tag{5.28}$$

The advantage of replacing a sum by an integral is that integrals are usually much easier to handle. Many more closed-form expressions are available for definite and indefinite integrals than for sums. We will discuss extensions of this principle of replacing sums by integrals further in section 5.3, when we present the Euler–Maclaurin summation formula. Usually, though, we do not need anything sophisticated, and the application of the principle to situations like that of (5.28) is easy to justify. If $a_n = g(n)$ for some function $g(x)$ of a real argument x, then

$$\left| g(n) - \int_n^{n+1} g(u)\,\mathrm{d}u \right| \leqslant \max_{n<u<n+1} |g(u) - g(n)| , \tag{5.29}$$

and so

$$\left| \sum_n g(n) - \int g(u)\,\mathrm{d}u \right| \leqslant \sum_n \max_{n\leqslant u\leqslant n+1} |g(u) - g(n)| , \tag{5.30}$$

where the integral is over $[a, b+1]$ if the sum is over $a \leqslant n \leqslant b$, $a, b \in \mathbb{Z}$. If $g(u)$ is continuously differentiable, then $|g(u) - g(n)| \leqslant \max_{n\leqslant v\leqslant n+1} |g'(v)|$ for $n \leqslant u \leqslant n+1$. This gives the estimate

$$\left| \sum_{n=a}^{b} g(n) - \int_a^{b+1} g(u)\,\mathrm{d}u \right| \leqslant \sum_{n=a}^{b} \max_{n\leqslant v\leqslant n+1} |g'(v)| . \tag{5.31}$$

Often one can find a simple explicit function $h(w)$ such that $|g'(v)| \leqslant h(w)$ for any v and w with $|v - w| \leqslant 1$, in which case eq. (5.31) can be replaced by

$$\left| \sum_{n=a}^{b} g(n) - \int_a^{b+1} g(u)\,\mathrm{d}u \right| \leqslant \int_a^{b+1} h(v)\,\mathrm{d}v . \tag{5.32}$$

For good estimates to be obtained from integral approximations to sums, it is usually necessary for individual terms to be small compared to the sum.

Example 5.3 (*Sum of* $\exp(-\alpha k^2)$). In the final stages of an asymptotic approximation one often encounters sums of the form

$$h(\alpha) = \sum_{k=-\infty}^{\infty} \exp(-\alpha k^2) , \qquad \alpha > 0 . \tag{5.33}$$

There is no closed form for the indefinite integral of $\exp(-\alpha u^2)$ (it is expressible in terms of the Gaussian error function only), but there is the famous evaluation of the definite integral

$$\int_{-\infty}^{\infty} \exp(-\alpha u^2)\,\mathrm{d}u = (\pi/\alpha)^{1/2} . \tag{5.34}$$

Thus it is natural to approximate $h(\alpha)$ by $(\pi/\alpha)^{1/2}$. If $g(u) = \exp(-\alpha u^2)$, then $g'(u) = -2\alpha u g(u)$, and so for $n \geqslant 0$,

$$\max_{n\leqslant v\leqslant n+1} |g'(v)| \leqslant 2\alpha(n+1)g(n) . \tag{5.35}$$

For the integral in eq. (5.30) to yield a good approximation to the sum we must show that the error term is smaller than the integral. The largest term in the sum occurs at $n = 0$ and equals 1. The error bound (5.35) that comes from approximating $g(0) = 1$ by the integral of $g(u)$ over $0 \leqslant u \leqslant 1$ is 2α. Therefore we cannot expect to obtain a good estimate unless $\alpha \to 0$. We find that

$$2\alpha(n+1)g(n) \leqslant 4\alpha u g(u/2) \quad \text{for } n \geqslant 1, \quad n \leqslant u \leqslant n+1 ,$$

so (integral approximation again!)

$$\begin{aligned} \textstyle\sum_{n=1}^{\infty} 2\alpha(n+1)g(n) &\leqslant 4\alpha \textstyle\int_1^{\infty} ug(u/2)\,du \\ &\leqslant 4\alpha \textstyle\int_0^{\infty} ug(u/2)\,du = (8\alpha)^{1/2} . \end{aligned} \tag{5.36}$$

Therefore, taking into account the error for $n = 0$ which was not included in the bound (5.36), we have

$$\begin{aligned} h(\alpha) &= \sum_{n=-\infty}^{\infty} \exp(-\alpha n^2) = \int_{-\infty}^{\infty} \exp(-\alpha u^2)\,du + O(\alpha^{1/2} + \alpha) \\ &= (\pi/\alpha)^{1/2} + O(\alpha^{1/2}) \quad \text{as } \alpha \to 0^+ . \end{aligned} \tag{5.37}$$

For this sum much more precise estimates are available, as will be shown in Example 5.9. For many purposes, though, (5.37) is sufficient. ☒

Example 5.3 showed how to use the basic tool of approximating a sum by an integral. Moreover, the estimate (5.37) that it provides is ubiquitous in asymptotic enumeration, since many approximations reduce to it. This is illustrated by the following example.

Example 5.4 (*Bell numbers* (cf. de Bruijn 1958)). The Bell number, $B(n)$, counts the partitions of an n-element set. It is given by (Comtet 1974):

$$B(n) = e^{-1} \sum_{k=1}^{\infty} \frac{k^n}{k!} . \tag{5.38}$$

In this sum no single term dominates. The ratio of the $(k+1)$st to the kth term is

$$\frac{(k+1)^n}{(k+1)!} \cdot \frac{k!}{k^n} = \frac{1}{k+1}\left(1 + \frac{1}{k}\right)^n . \tag{5.39}$$

As k increases, this ratio strictly decreases. We search for the point where it is about 1. For $k \geqslant 2$,

$$\left(1 + \frac{1}{k}\right)^n = \exp\left(n \log\left(1 + \frac{1}{k}\right)\right) = \exp(n/k + O(n/k^2)) , \tag{5.40}$$

so the ratio is close to 1 for n/k close to $\log(k+1)$. We choose k_0 to be the closest integer to w, the solution to

$$n = w \log(w+1) . \tag{5.41}$$

For $k = k_0 + j$, $1 \leqslant j \leqslant k_0/2$, we find, since $\log(1 + i/k_0) = i/k_0 - i^2/(2k_0^2) + \mathrm{O}(i^3/k_0^3)$,

$$\begin{aligned}\frac{k^n}{k!} &= \frac{k_0^n}{k_0!}\,\frac{(1+j/k_0)^n}{k_0^j\Pi_{i=1}^j(1+i/k_0)}\\ &= \frac{k_0^n}{k_0!}\exp\left(jn/k_0 - j\log k_0 - j^2(n+k_0)/(2k_0^2) + \mathrm{O}(nj^3/k_0^3 + j/k_0)\right)\ .\end{aligned} \tag{5.42}$$

The same estimate applies for $-k_0/2 \leqslant j \leqslant 0$. The term $jn/k_0 - j\log k_0$ is small, since $|k_0 - w| \leqslant 1/2$ and w satisfies (5.41). We find

$$\begin{aligned}n/k_0 - \log k_0 &= n/w - \log(w+1) + \mathrm{O}(n/w^2 + 1/w)\\ &= \mathrm{O}(n/w^2 + 1/w)\ .\end{aligned} \tag{5.43}$$

By (5.41), $w \sim n/\log n$ as $n \to \infty$. We now further restrict j to $|j| \leqslant n^{1/2}\log n$. Then (5.42) and (5.43) yield

$$\frac{k^n}{k!} = \frac{k_0^n}{k_0!}\exp\left(-j^2(n+k_0)/(2k_0^2) + \mathrm{O}((\log n)^6 n^{-1/2})\right)\ . \tag{5.44}$$

Approximating the sum by an integral, as in Example 5.3, shows that

$$\sum_{\substack{k\\ |j|\leqslant n^{1/2}\log n}} \frac{k^n}{k!} = \frac{k_0^n}{k_0!}k_0(2\pi)^{1/2}(n+k_0)^{-1/2}(1 + \mathrm{O}((\log n)^6 n^{-1/2}))\ . \tag{5.45}$$

(An easy way to obtain this is to apply the estimate of Example 5.3 to the sum from $-\infty$ to ∞, and show that the range $|j| > n^{1/2}\log n$ contributes little.) To estimate the contribution of the remaining summands, with $|j| > n^{1/2}\log n$, we observe that the ratio of successive terms is $\leqslant 1$, so the range $1 \leqslant k \leqslant k_0 - \lfloor n^{1/2}\log n\rfloor$ contributes at most k_0 (the number of terms) times the largest term, which arises for $k = k_0 - \lfloor n^{1/2}\log n\rfloor$. By (5.44), this largest term is

$$\mathrm{O}(k_0^n(k_0!)^{-1}\exp(-(\log n)^3))\ .$$

For $k \geqslant k_1 \geqslant k_0 + \lfloor n^{1/2}\log n\rfloor$, we find that the ratio of the $(k+1)$st to the kth term is, for large n,

$$\begin{aligned}\leqslant \frac{1}{k_1+1}\left(1+\frac{1}{k_1}\right)^n &= \exp(n/k_1 - \log(k_1+1) - n/(2k_1^2) + \mathrm{O}(n/k_1^3))\\ &\leqslant \exp(-(k_1-k_0)n/k_1^2 + \mathrm{O}(n/k_1^3))\\ &\leqslant \exp(-2n^{-1/2})\ \leqslant\ 1 - n^{-1/2}\ ,\end{aligned} \tag{5.46}$$

and so the sum of these terms, for $k_1 \leqslant k < \infty$, is bounded above by $n^{1/2}$ times the term for $k = k_1$. Therefore the estimate on the right-hand side of (5.45) applies even when we sum on all k, $1 \leqslant k < \infty$.

To obtain an estimate for $B(n)$, it remains only to estimate $k_0^n/k_0!$. To do this, we apply Stirling's formula and use the property that $|k_0 - w| \leqslant 1/2$ to deduce that

$$B(n) \sim (\log w)^{1/2} w^{n-w} \mathrm{e}^w \quad \text{as } n \to \infty , \tag{5.47}$$

where w is given by (5.41).

There is no explicit formula for w in terms of n, and substituting various asymptotic approximations to w, such as

$$w = \frac{n}{\log n} + \mathrm{O}\left(\frac{n}{(\log n)^2}\right) \tag{5.48}$$

(see Example 5.10) yields large error terms in (5.47), so for accuracy it is usually better to use (5.47) as it is. There are other approximations to $B(n)$ in the literature (see, for example, Bender 1974, de Bruijn 1958). They differ slightly from (5.47) because they estimate $B(n)$ in terms of roots of equations other than (5.41).

Other methods of estimating $B(n)$ are presented in Examples 12.9 and 12.13.

☒

5.2. Alternating sums and the principle of inclusion–exclusion

At the beginning of section 5, the reader was advised in general to search for identities and transformations when dealing with general sums. This advice is even more important when dealing with sums of terms that have alternating or irregularly changing coefficients. Finding the largest term is of little help when there is substantial cancellation among terms. Several general approaches for dealing with this difficulty will be presented later. Generating-function methods for dealing with complicated sums are discussed in section 6. Contour-integration methods for alternating sums are mentioned in section 10.3. The summation formulas of the next section can sometimes be used to estimate sums with regularly varying coefficients as well. In this section we present some basic elementary techniques that are often sufficient.

Sometimes it is possible to obtain estimates of sums with positive and negative summands by approximating separately the sums of the positive and of the negative summands. Methods of the preceding section or of the next section are useful in such situations. However, this approach is to be avoided as much as possible, because it often requires extremely precise estimates of the two sums to obtain even rough bounds on the desired sums. One method that often works and is much simpler consists of a simple pairing of adjacent positive and negative terms.

Example 5.5 (*Alternating sum of square roots*). Let

$$S_n = \sum_{k=1}^{n} (-1)^k k^{1/2} . \tag{5.49}$$

We have

$$(2m)^{1/2} - (2m-1)^{1/2} = (2m)^{1/2}\left\{1 - \left(1 - \frac{1}{2m}\right)^{1/2}\right\}$$

$$= (2m)^{1/2}\left\{1 - \left(1 - \frac{1 - \frac{1}{4m} +}{4m} + O(m^{-2})\right)\right\} \tag{5.50}$$

$$= (8m)^{-1/2} + O(m^{-3/2}) \, ,$$

so

$$\sum_{k=1}^{2\lfloor n/2\rfloor} (-1)^k k^{1/2} = \sum_{m=1}^{\lfloor n/2\rfloor} (8m)^{-1/2} + O(1)$$

$$= n^{1/2}/2 + O(1) \, . \tag{5.51}$$

Hence

$$S_n = \begin{cases} n^{1/2}/2 + O(1) & \text{if } n \text{ is even,} \\ -n^{1/2}/2 + O(1) & \text{if } n \text{ is odd.} \end{cases} \tag{5.52}$$

☒

In Example 5.5, the sums of the positive terms and of the negative terms can easily be estimated accurately (for example, by using the Euler–Maclaurin formula of the next section) to obtain (5.52). In other cases, though, the cancellation is too extensive for such an approach to work. This is especially true for sums arising from the principle of inclusion–exclusion.

Suppose that X is some set of objects and P is a set of properties. For $R \subseteq P$, let $N_=(R)$ be the number of objects in X that have exactly the properties in R and none of the properties in $P \setminus R$. We let $N_\geqslant(R)$ denote the number of objects in X that have all the properties in R and possibly some of those in $P \setminus R$. The principle of inclusion–exclusion says that

$$N_=(R) = \sum_{R \subseteq Q \subseteq P} (-1)^{|Q \setminus R|} N_\geqslant(Q) \, . \tag{5.53}$$

[This is a basic version of the principle. For more general results, proofs, and references, see Comtet (1974), Goulden and Jackson (1983), and Stanley (1986).]

Example 5.6 (*Derangements of n letters*). Let X be the set of permutations of n letters, and suppose that P_i, $1 \leqslant i \leqslant n$, is the property that the ith letter is fixed by a permutation, and $P = \{P_1, \ldots, P_n\}$. Then d_n, the number of derangements of n letters, equals $N_=(\emptyset)$, where $\emptyset$ is the empty set, and so by (5.53)

$$d_n = \sum_{Q \subseteq P} (-1)^{|Q|} N_\geqslant(Q) \, . \tag{5.54}$$

However, $N_\geqslant(Q)$ is just the number of permutations that leave all letters specified by Q fixed, and thus

$$d_n = \sum_{Q \subseteq P} (-1)^{|Q|} (n - |Q|)!$$

$$= \sum_{k=0}^{n} (-1)^k (n-k)! \binom{n}{k} = \sum_{k=0}^{n} (-1)^k \frac{n!}{k!} \, , \tag{5.55}$$

which is eq. (1.1). ☒

The formula (1.1) for derangements is easy to use because the terms decrease rapidly. Moreover, this formula is exceptionally simple, largely because $N_{\geqslant}(Q)$ depends only on $|Q|$. In general, the inclusion–exclusion principle produces complicated sums that are hard to estimate. A frequently helpful tool is provided by the *Bonferroni inequalities* (Comtet 1974, Stanley 1986). One form of these inequalities is that for any integer $m \geqslant 0$,

$$N_{=}(R) \geqslant \sum_{\substack{Q \\ R\subseteq Q\subseteq P \\ |Q\setminus R|\leqslant 2m}} (-1)^{|Q\setminus R|} N_{\geqslant}(Q) \tag{5.56}$$

and

$$N_{=}(R) \leqslant \sum_{\substack{Q \\ R\subseteq Q\subseteq P \\ |Q\setminus R|\leqslant 2m+1}} (-1)^{|Q\setminus R|} N_{\geqslant}(Q)\,. \tag{5.57}$$

Thus in general

$$\Bigg|N_{=}(R) - \sum_{\substack{Q \\ R\subseteq Q\subseteq P \\ |Q\setminus R|\leqslant k}} (-1)^{|Q\setminus R|} N_{\geqslant}(Q)\Bigg| \leqslant \sum_{\substack{Q \\ R\subseteq Q\subseteq P \\ |Q\setminus R|\leqslant k+1}} N_{\geqslant}(Q)\,. \tag{5.58}$$

These inequalities are frequently applied for $n = |X|$ increasing. Typically one chooses k that increases much more slowly than n, so that the individual terms $N_{\geqslant}(Q)$ in (5.58) can be estimated asymptotically, as the interactions of the different properties counted by $N_{\geqslant}(Q)$ are not too complicated to estimate. Bender (1974) presents some useful general principles to be used in such estimates (especially the asymptotically Poisson distribution that tends to occur when the method is successful). We present an adaptation of an example from Bender (1974).

Example 5.7 (*Balls and cells*). Given n labeled cells and m labeled balls, let $a_h(m,n)$ be the number of ways to place the balls into cells so that exactly h of the cells are empty. We consider h fixed. Let X be the ways of placing the balls into the cells (n^m in total), and $P = \{P_1, \ldots, P_n\}$, where P_i is the property that the ith cell is empty. If $R = \{P_1, \ldots, P_h\}$, then $a_h(m,n) = \binom{n}{h} N_{=}(R)$. Now

$$N_{\geqslant}(Q) = (n - |Q|)^m\,, \tag{5.59}$$

so

$$\begin{aligned}\sum_{\substack{Q \\ R\subseteq Q\subseteq P \\ |Q\setminus R|=t}} N_{\geqslant}(Q) &= \binom{n-h}{t}(n-h-t)^m \\ &= n^m\,\mathrm{e}^{-mh/n}(n\,\mathrm{e}^{-m/n})^t (t!)^{-1}(1 + \mathrm{O}((t^2+1)mn^{-2} + (t^2+1)n^{-1}))\,,\end{aligned} \tag{5.60}$$

provided $t^2 \leqslant n$ and $mt^2n^{-2} \leqslant 1$, say. In the range $0 \leqslant t \leqslant \log n$, $n \log n \leqslant m \leqslant n^2(\log n)^{-3}$, we find that the right-hand side of (5.60) is

$$n^m\,\mathrm{e}^{-mh/n}(n\,\mathrm{e}^{-m/n})^t (t!)^{-1}(1 + \mathrm{O}(mn^{-2}(\log n)^2))\,.$$

We now apply (5.58) with $k = \lfloor \log n \rfloor$, and obtain

$$\begin{aligned} a_h(m,n) &= \tbinom{n}{h} N_=(R) \sim \tbinom{n}{h} n^m \exp(-mh/n - n\,\mathrm{e}^{-m/n}) \\ &\sim n^m (h!)^{-1} (n\,\mathrm{e}^{-m/n})^h \exp(-n\,\mathrm{e}^{-m/n}) \end{aligned} \tag{5.61}$$

as $m, n \to \infty$, provided $n \log n \leqslant m \leqslant n^2 (\log n)^{-3}$. Since $a_h(m,n) n^{-m}$ is the probability that there are exactly h empty cells, the relation (5.61) (which we have established only for fixed h) shows that this probability is asymptotically distributed like a Poisson random variable with parameter $n \exp(-m/n)$.

Many additional results on random distributions of balls into cells, and references to the extensive literature on this subject can be found in Kolchin et al. (1978).

⊠

Bonferroni inequalities include other methods for estimating $N_=(R)$ by linear combinations of the $N_\geqslant(Q)$. Recent approaches and references (phrased in probabilistic terms) can be found in Galambos (1977). For bivariate Bonferroni inequalities (where one asks for the probability that at least one of two sets of events occurs) see Galambos and Xu (1993) and Lee (1992).

The Chen–Stein method (Chen 1975) is a powerful technique that is often used in place of the principle of inclusion–exclusion, especially in probabilistic literature. Recent references are Arratia et al. (1990a) and Barbour et al. (1992).

5.3. *Euler–Maclaurin and Poisson summation formulas*

The introduction to section 5 showed that sums can be successfully approximated by integrals if the summands are all small compared to the total sum and vary smoothly as functions of the summation index. The approximation (5.29), though crude, is useful in a wide variety of cases. Sometimes, though, more accurate approximations are needed. An obvious way is to improve the bound (5.29). If $g(x)$ is really smooth, we can expect that the difference

$$a_n - \int_n^{n+1} g(u)\,\mathrm{d}u$$

will vary in a regular way with n. This is indeed the case, and it is exploited by the Euler–Maclaurin summation formula. It can be found in many books, such as de Bruijn (1958), Graham et al. (1989), Abramowitz and Stegun (1970), and Nörlund (1924). There are many formulations, but they do not differ much.

Euler–Maclaurin summation formula

Suppose that $g(x)$ has $2m$ continuous derivatives in $[a,b]$, $a, b \in \mathbb{Z}$. Then

$$\begin{aligned} \sum_{k=a}^{b} g(k) = \int_a^b g(x)\,\mathrm{d}x + \sum_{r=1}^{m} \frac{B_{2r}}{(2r)!} \left\{ g^{(2r-1)}(b) - g^{(2r-1)}(a) \right\} \\ + \frac{1}{2}\{g(a) + g(b)\} + R_m\ , \end{aligned} \tag{5.62}$$

where

$$R_m = -\int_a^b g^{(2m)}(x)\frac{B_{2m}(x-\lfloor x\rfloor)}{(2m)!}\,\mathrm{d}x\ , \tag{5.63}$$

and so

$$|R_m| \leqslant \int_a^b |g^{(2m)}(x)|\frac{|B_{2m}(x-\lfloor x\rfloor)|}{(2m)!}\,\mathrm{d}x\ . \tag{5.64}$$

In the above formulas, the $B_n(x)$ denote the Bernoulli polynomials, defined by

$$\frac{z\,\mathrm{e}^{xz}}{\mathrm{e}^z-1} = \sum_{n=0}^{\infty} B_n(x)\frac{z^n}{n!}\ . \tag{5.65}$$

The B_n are the Bernoulli numbers, defined by

$$\frac{z}{\mathrm{e}^z-1} = \sum_{n=0}^{\infty} B_n\frac{z^n}{n!}\ , \tag{5.66}$$

so that $B_n = B_n(0)$, and

$$\begin{aligned} &B_0 = 1\ ,\quad B_1 = -1/2\ ,\quad B_2 = 1/6\ ,\\ &B_3 = B_5 = B_7 = \cdots = 0\ ,\\ &B_4 = -1/30\ ,\quad B_6 = 1/42\ ,\quad B_8 = -1/30,\ \ldots\ . \end{aligned} \tag{5.67}$$

It is known that

$$|B_{2m}(x-\lfloor x\rfloor)| \leqslant |B_{2m}|\ , \tag{5.68}$$

so we can simplify (5.64) to

$$|R_m| \leqslant |B_{2m}|((2m)!)^{-1}\int_a^b |g^{(2m)}(x)|\ \mathrm{d}x\ . \tag{5.69}$$

There are many applications of the Euler–Maclaurin formula. One of the most frequently cited ones is to estimate factorials.

Example 5.8 (*Stirling's formula*). We transform the product in the definition of $n!$ into a sum by taking logarithms, and find that for $g(x) = \log x$ and $m = 1$ we have

$$\log n! = \sum_{k=1}^{n}\log k = \int_1^n (\log x)\,\mathrm{d}x + \frac{1}{2}\log n + \frac{1}{2}B_2\left\{\frac{1}{n}-1\right\} + R_1\ , \tag{5.70}$$

where

$$R_1 = \int_1^n \frac{B_2(x-\lfloor x\rfloor)}{2x^2}\,\mathrm{d}x = C + \mathrm{O}(n^{-1}) \tag{5.71}$$

for

$$C = \int_1^\infty \frac{B_2(x - \lfloor x \rfloor)}{2x^2}\,\mathrm{d}x\,. \tag{5.72}$$

Therefore

$$\log n! = n \log n - n + \frac{1}{2}\log n + C + 13/12 + \mathrm{O}(n^{-1})\,, \tag{5.73}$$

which gives

$$n! \sim C' n^{1/2} n^n \mathrm{e}^{-n} \quad \text{as } n \to \infty\,. \tag{5.74}$$

To obtain Stirling's formula (4.1), we need to show that $C' = (2\pi)^{1/2}$. This can be done in several ways (cf. de Bruijn 1958). In Examples 12.1, 12.7, and 12.9 we will see other methods of deriving (4.1). ⊠

There is no requirement that the function $g(x)$ in the Euler–Maclaurin formula be positive. That was not even needed for the crude approximation of a sum by an integral given in section 5.0. The function $g(x)$ can even take complex values. [After all, eq. (5.62) is an identity!] However, in most applications this formula is used to derive an asymptotic estimate with a small error term. For that, some high-order derivatives have to be small, which means that $g(x)$ cannot change sign too rapidly. In particular, the Euler–Maclaurin formula usually is not very useful when the $g(k)$ alternate in sign. In those cases one can sometimes use the differencing trick (cf. Example 5.5) and apply the Euler–Maclaurin formula to $h(k) = g(2k) + g(2k + 1)$. There is also Boole's summation formula for alternating sums that can be applied. [See chapter 2, section 3 and chapter 6, section 6 of Nörlund (1924), for example.] Generalizations to other periodic patterns in the coefficients have been derived by Berndt and Schoenfeld (1975/76).

The bounds for the error term R_m in the Euler–Maclaurin formula that were stated above can often be improved by using special properties of the function $g(x)$. For example, when $g(x)$ is analytic in x, there are contour integrals for R_m that sometimes give good estimates (cf. Olver 1974).

The Poisson summation formula states that

$$\sum_{n=-\infty}^{\infty} f(n + a) = \sum_{m=-\infty}^{\infty} \exp(2\pi i m a) \int_{-\infty}^{\infty} f(y) \exp(-2\pi i m y)\,\mathrm{d}y \tag{5.75}$$

for "nice" functions $f(x)$. The functions for which (5.75) holds include all continuous $f(x)$ for which $\int |f(x)|\,\mathrm{d}x < \infty$, which are of bounded variation, and for which $\sum_n f(n + a)$ converges for all a. For weaker conditions that ensure validity of (5.75), we refer to de Bruijn (1958) and Titchmarsh (1948). The Poisson summation formula often converts a slowly convergent sum into a rapidly convergent one. Generally it is not as widely applicable as the Euler–Maclaurin formula as it requires extreme regularity for the Fourier coefficients to decrease rapidly. On the other hand, it can be applied in some situations that are not covered by the Euler–Maclaurin formula, including some where the coefficients vary in sign.

Example 5.9 (*Sum of* $\exp(-\alpha k^2)$). We consider again the function $h(\alpha)$ of Example 5.3. We let $f(x) = \exp(-\alpha x^2)$, $a = 0$. Equation (5.15) then gives

$$h(\alpha) = \sum_{n=-\infty}^{\infty} \exp(-\alpha n^2) = (\pi/\alpha)^{1/2} \sum_{m=-\infty}^{\infty} \exp(-\pi^2 m^2/\alpha) . \qquad (5.76)$$

This is an identity, and the sum on the right-hand side above converges rapidly for small α. Many applications require the evaluation of the sum on the left in which α tends to 0. Equation (5.76) offers a method of converting a slowly convergent sum into a tractable one, whose asymptotic behavior is explicit. ⊠

5.4. *Bootstrapping and other basic methods*

Bootstrapping is a useful technique that uses asymptotic information to obtain improved estimates. Usually we start with some rough bounds, and by combining them with the relations defining the function or sequence that we are studying, we obtain better bounds.

Example 5.10 (*Approximation of Bell numbers*). Example 5.4 obtained the asymptotics of the Bell numbers B_n, but only in terms of w, the solution to eq. (5.41). We now show how to obtain asymptotic expansions for w. As n increases, so does w. Therefore $\log(w + 1)$ also increases, and so $w < n$ for large n. Thus

$$n = w \log(w + 1) < w \log(n + 1) ,$$

and so

$$\mathrm{n}(\log(n + 1))^{-1} < w < n . \qquad (5.77)$$

Therefore

$$\log(w + 1) = \log n + \mathrm{O}(\log \log n) , \qquad (5.78)$$

and so

$$w = \frac{n}{\log(w+1)} = \frac{n}{\log n} + \mathrm{O}\left(\frac{n \log \log n}{(\log n)^2}\right) . \qquad (5.79)$$

To go further, note that by (5.79),

$$\begin{aligned} \log(w + 1) &= \log\left(\frac{n}{\log n}\left(1 + \mathrm{O}\left(\frac{\log \log n}{\log n}\right)\right)\right) \\ &= \log n - \log \log n + \mathrm{O}((\log \log n)(\log n)^{-1}) , \end{aligned} \qquad (5.80)$$

and so by applying this estimate in eq. (5.41), we obtain

$$w = \frac{n}{\log n} + \frac{n \log \log n}{(\log n)^2} + \frac{n(\log \log n)^2}{(\log n)^3} + \mathrm{O}\left(\frac{n \log \log n}{(\log n)^3}\right) . \qquad (5.81)$$

This procedure can be iterated indefinitely to obtain expansions for w with error terms $\mathrm{O}(n(\log n)^{-\alpha})$ for as large a value of α as desired. ⊠

In the above example, w can also be estimated by other methods, such as the Lagrange–Bürmann inversion formula (cf. Example 6.7). However, the bootstrapping method is much more widely applicable and easy to apply. It will be used several times later in this chapter.

5.5. *Estimation of integrals*

In some of the examples in the preceding sections integrals were used to approximate sums. The integrals themselves were always easy to evaluate. That is true in most asymptotic enumeration problems, but there do occur situations where the integrals are more complicated. Often the hard integrals arc of thc form

$$f(x) = \int_{\alpha}^{\beta} g(t)\exp(xh(t))\,\mathrm{d}t\ , \tag{5.82}$$

and it is necessary to estimate the behavior of $f(x)$ as $x \to \infty$, with the functions $g(t)$, $h(t)$ and the limits of integration α and β held fixed. There is a substantial theory of such integrals, and good references are Bleistein and Handelsman (1975), de Bruijn (1958), Erdélyi (1956), and Olver (1974). The basic technique is usually referred to as Laplace's method, and consists of approximating the integrand by simpler functions near its maxima. This approach is similar to the one that is discussed at length in section 5.1 for estimating sums. The contributions of the approximations are then evaluated, and it is shown that the remaining ranges of integration, away from the maxima, contribute a negligible amount. By breaking up the interval of integration we can write the integral (5.82) as a sum of several integrals of the same type, with the property that there is a unique maximum of the integrand and that it occurs at one of the endpoints. When $\alpha > 0$, the maximum of the integrand occurs for large x at the maximum of $h(t)$ (except in rare cases where $g(t) = 0$ for that t for which $h(t)$ is maximized). Suppose that the maximum occurs at $t = \alpha > 0$. It often happens that

$$h(t) = h(\alpha) - c(t-\alpha)^2 + \mathrm{O}(|t-\alpha|^3) \tag{5.83}$$

for $\alpha \leqslant t \leqslant \beta$ and $c = -h''(\alpha)/2 > 0$, and then one obtains the approximation

$$f(x) \sim g(\alpha)\exp(xh(\alpha))[-\pi/(4xh''(\alpha))]^{1/2} \quad \text{as } x \to \infty\ , \tag{5.84}$$

provided $g(\alpha) \neq 0$. For precise statements of even more general and rigorous results, see for example chapter 3, section 7 of Olver (1974). Those results cover functions $h(t)$ that behave near $t = \alpha$ like $h(\alpha) - c(t-\alpha)^{\mu}$ for any $\mu > 0$.

When the integral is highly oscillatory, as happens when $h(t) = \mathrm{i}u(t)$ for a real-valued function $u(t)$, still other techniques (such as the stationary phase method), are used. We will not present them here, and refer to the references cited above for descriptions and applications. In section 12.1 we will discuss the saddle point method, which is related to both Laplace's method and the stationary phase method.

Laplace integrals

$$F(x) = \int_0^{\infty} f(t)\exp(-xt)\,\mathrm{d}t \tag{5.85}$$

can often be approximated by integration by parts. We have (under suitable conditions on $f(t)$)

$$F(x) = x^{-1}f(0) + x^{-1}\int_0^\infty f'(t)\exp(-xt)\,\mathrm{d}t$$
$$= x^{-1}f(0) + x^{-2}f'(0) + x^{-2}\int_0^\infty f''(t)\exp(-xt)\,\mathrm{d}t\ , \tag{5.86}$$

and so on. There are general results, usually associated with the name of Watson's Lemma, for deriving such expansions. For references, see Erdélyi (1956) or Olver (1974).

6. Generating functions

6.1. A brief overview

Generating functions are a wonderfully powerful and versatile tool, and most asymptotic estimates are derived from them. The most common ones in combinatorial enumeration are the ordinary and exponential generating functions. If $a_0, a_1, \ldots,$ is any sequence of real or complex numbers, the *ordinary generating function* is

$$f(z) = \sum_{n=0}^{\infty} a_n z^n\ , \tag{6.1}$$

while the *exponential generating function* is

$$f(z) = \sum_{n=0}^{\infty} \frac{a_n z^n}{n!}\ . \tag{6.2}$$

Doubly-indexed arrays, for example $a_{n,k}$, $0 \leqslant n < \infty$, $0 \leqslant k \leqslant n$, are encoded as two-variable generating functions. Depending on the array, sometimes one uses

$$f(x,y) = \sum_{n=0}^{\infty}\sum_{k=0}^{n} a_{n,k}x^k y^n\ , \tag{6.3}$$

and sometimes other forms that might even mix ordinary and exponential types, as in

$$f(x,y) = \sum_{n=0}^{\infty}\frac{y^n}{n!}\sum_{k=0}^{n} a_{n,k}x^k\ . \tag{6.4}$$

For example, the Stirling numbers of the first kind, $s(n,k)$ ($(-1)^{n+k}s(n,k)$ is the number of permutations on n letters with k cycles) have the generating function (see pp. 50, 212–213, and 234–235 in Comtet 1974):

$$1 + \sum_{n=1}^{\infty}\frac{y^n}{n!}\sum_{k=1}^{n} s(n,k)x^k = (1+y)^x\ . \tag{6.5}$$

In general, a generating function is just a formal power series, and questions of convergence do not arise in the definition. However, some of the main applications of generating functions in asymptotic enumeration do rely on analyticity or other convergence properties of those functions, and there the domain of convergence is important.

A generating function is just another form for the sequence that defines it. There are many reasons for using it. One is that even for complicated sequences, generating functions are frequently simple. This might not be obvious for the partition function $p(n)$, which has the ordinary generating function

$$f(z) = \sum_{n=0}^{\infty} p(n) z^n = \prod_{k=1}^{\infty} (1 - z^k)^{-1} . \tag{6.6}$$

The sequence $p(n)$, which is complicated, is encoded here as an infinite product. The terms in the product are simple and vary in a regular way with the index, but it is not clear at first what is gained by this representation. In other cases, though, the advantages of generating functions are clearer. For example, the exponential generating function for derangements [eq. (1.1) and Example 5.6] is

$$\begin{aligned} f(z) &= \sum_{n=0}^{\infty} \frac{d_n}{n!} z^n = \sum_{n=0}^{\infty} \frac{z^n}{n!} \sum_{n=0}^{n} (-1)^k \frac{n!}{k!} \\ &= \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \sum_{n=k}^{\infty} z^n = \frac{e^{-z}}{1-z} , \end{aligned} \tag{6.7}$$

which is extremely compact.

Reasons for using generating functions go far beyond simplicity. The one that matters most for this chapter in that generating functions can be used to obtain information about the asymptotic behavior of sequences they encode, information that often cannot be obtained in any other way, or not as easily. Methods such as those of section 10.2 can be used to obtain immediately from eq. (6.7) the asymptotic estimate $d_n \sim e^{-1} n!$ as $n \to \infty$. This estimate can also be derived easily by elementary methods from eq. (1.1), so here the generating function is not essential. In other cases, though, such as that of the partition function $p(n)$, all the sharp estimates, such as that of Hardy and Ramanujan given in eq. (1.5), are derived by exploiting the properties of the generating function. If there is any main theme to this chapter, it is that generating functions are usually the easiest, most versatile, and most powerful way to study asymptotic behavior of sequences. Especially when the generating function is analytic, its behavior at the dominant singularities (a term that will be defined in section 10) determines the asymptotics of the sequence. When the generating function is simple, and often even when it is not simple, the contribution of the dominant singularity can often be determined easily, although the sequence itself is complicated.

There are many applications of generating functions, some related to asymptotic questions. Averages can often be studied using generating functions. Suppose, for example, that $a_{n,k}$, $0 \leqslant k \leqslant n$, $0 \leqslant n < \infty$, is the number of objects in some class of size n, which have weight k (for some definition of size and weight), and that we know, either explicitly or implicitly, the generating function $f(x, y)$ of $a_{n,k}$ given by (6.4). Then

$$g(y) = f(1, y) = \sum_{n=0}^{\infty} \frac{y^n}{n!} \sum_{k=0}^{n} a_{n,k} \tag{6.8}$$

is the exponential generating function of the number of objects of size n, while

$$h(y) = \frac{\partial}{\partial x} f(x, y)\bigg|_{x=1} = \sum_{n=0}^{\infty} \frac{y^n}{n!} \sum_{k=0}^{n} k a_{n,k} \tag{6.9}$$

is the exponential generating function of the sum of the weights of objects of size n. Therefore the average weight of an object of size n is

$$\frac{[y^n]h(y)}{[y^n]g(y)} . \tag{6.10}$$

The wide applicability and power of generating functions come primarily from the structured way in which most enumeration problems arise. Usually the class of objects to be counted is derived from simpler objects through basic composition rules. When the generating functions are chosen to reflect appropriately the classes of objects and composition rules, the final generating function is derivable in a simple way from those of the basic objects. Suppose, for example, that each object of size n in class C can be decomposed uniquely into a pair of objects of sizes k and $n - k$ (for some k) from classes A and B, and each pair corresponds to an object in C. Then c_n, the number of objects of size n in C, is given by the convolution

$$c_n = \sum_{k=0}^{n} a_k b_{n-k} , \tag{6.11}$$

(where a_k is the number of objects of size k in A, etc.). Hence if $A(z) = \sum a_n z^n$, $B(z) = \sum b_n z^n$, $C(z) = \sum c_n z^n$ are the ordinary generating functions, then

$$C(z) = A(z)B(z) . \tag{6.12}$$

Thus ordered pairing of objects corresponds to multiplication of ordinary generating functions. If $A(z) = \sum a_n z^n$ and

$$b_n = \sum_{k=0}^{n} a_k ,$$

then $B(z) = \sum b_n z^n$ is given by

$$B(z) = \frac{A(z)}{1 - z} , \tag{6.13}$$

so that the ordinary generating function of cumulative sums of coefficients is obtained by dividing by $1 - z$. There are many more such general correspondences between operations on combinatorial objects and on the corresponding generating functions. They are present, implicitly or explicitly, in most books that cover combinatorial enumeration, such as Comtet (1974), Goulden and Jackson (1983), Stanley (1986), and Wilf (1990). The most systematic approach to developing and using general rules of this type has been carried out by Flajolet et al. (1991b). They develop ways to see immediately (cf. Flajolet and Odlyzko 1990a) that if we consider mappings of a set of n labeled elements to itself, so that all n^n distinct mappings are considered equally likely, then the generating function for the longest path length is given by

$$f(z) = \sum_{k=0}^{\infty} \left(\frac{1}{1 - t(z)} - e^{v_k(z)} \right) , \tag{6.14}$$

where

$$v_k(z) = t_{k-1}(z) + \frac{1}{2} t_{k-2}(z)^2 + \cdots + \frac{1}{k} t_0(k)^k , \tag{6.15}$$

with

$$t_0(z) = z , \quad t_{h+1}(z) = z \exp(t_h(z)) , \tag{6.16}$$

and $t(z) = \lim_{h \to \infty} t_h(z)$ (in the sense of formal power series, so convergence is that of coefficients). Furthermore, as is mentioned in section 17, many of these rules for composition of objects and generating functions can be implemented algorithmically, automating some of the chores of applying them.

We illustrate some of the basic generating function techniques by deriving the generating function for rooted labeled trees, which will occur later in Examples 6.6 and 10.14. (The rooted unlabeled trees, with generating function given by (1.8), are harder.)

Example 6.1 (*Rooted labeled trees*). Let t_n be the number of rooted labeled trees on n vertices, so that $t_1 = 1$, $t_2 = 2$, $t_3 = 9$. (It will be shown in Example 6.6 that $t_n = n^{n-1}$.) Let

$$t(z) = \sum_{n=1}^{\infty} t_n \frac{z^n}{n!} \tag{6.17}$$

be the exponential generating function. If we remove the root of a rooted labeled tree with n vertices, we are left with $k \geqslant 0$ rooted labeled trees that contain a total of $n - 1$ vertices. The total number of ways of arranging an ordered selection of k rooted trees with a total of $n - 1$ vertices is

$$[z^{n-1}] t(z)^k .$$

Since the order of the trees does not matter, we have

$$\frac{1}{k!} [z^{n-1}] t(z)^k$$

different trees of size n that have exactly k subtrees, and so

$$t_n = \sum_{k=0}^{\infty} \frac{1}{k!}[z^{n-1}]t(z)^k$$

$$= [z^{n-1}] \sum_{k=0}^{\infty} t(z)^k/k! = [z^n]z\exp(t(z)) , \tag{6.18}$$

which gives

$$t(z) = z\exp(t(z)) . \tag{6.19}$$

As an aside, the function $t_h(z)$ of eq. (6.16) is the exponential generating function of rooted labeled trees of height $\leqslant h$. ⊠

The key to the successful use of generating functions is to use a generating function that is of the appropriate form for the problem at hand. There is no simple rule that describes what generating function to use, and sometimes two are used simultaneously. In combinatorics and analysis of algorithms, the most useful forms are the ordinary and exponential generating functions, which reflects how the classes of objects that are studied are constructed. Sometimes other forms are used, such as the double exponential form

$$f(z) = \sum_{n=0}^{\infty} \frac{a_n z^n}{(n!)^2} \tag{6.20}$$

that occurs in section 7, or the Newton series

$$f(z) = \sum_{n=0}^{\infty} a_n z(z-1)\cdots(z-n+1) . \tag{6.21}$$

Also frequently encountered are various q-analog generating functions, such as the Eulerian

$$f(z) = \sum_{n=1}^{\infty} \frac{a_n z^n}{(1-q)(1-q^2)\cdots(1-q^n)} . \tag{6.22}$$

In multiplicative number theory, the most common are Dirichlet series

$$f(z) = \sum_{n=1}^{\infty} a_n n^{-z} , \tag{6.23}$$

which reflect the multiplicative structure of the integers. If a_n is a multiplicative function (so that $a_{mn} = a_m a_n$ for all relatively prime positive integers m and n) then the function (6.23) has an Euler-product representation

$$f(z) = \prod_p (1 + a_p p^{-z} + a_{p^2} p^{-2z} + \cdots) , \tag{6.24}$$

where p runs over the primes. This allows new tools to be used to study $f(z)$ and through it a_n. Additive problems in combinatorics and number theory often are handled using functions such as

$$f(z) = \sum_{n=1}^{\infty} z^{a_k} , \tag{6.25}$$

where $0 \leqslant a_1 < a_2 < \cdots$ is a sequence of integers. Addition of two such sequences then corresponds to a multiplication of the generating functions of the form (6.25).

We next mention the "snake-oil" method. This is the name given by Wilf (1990) to the use of generating functions for proving identities, and comes from the surprising power of this technique. The typical application is to evaluation of sequences given by sums of the type

$$a_n = \sum_k b_{n,k} . \tag{6.26}$$

The standard procedure is to form a generating function of the a_n and manipulate it through interchanges of summation and other tricks to obtain the final answer. The generating function can be ordinary, exponential, or (less commonly) of another type, depending on what gives the best results. We show a simple application of this principle that exhibits the main features of the method.

Example 6.2 (*A binomial coefficient sum,* Wilf 1990). Let

$$a_n = \sum_{k=0}^{n} \binom{n+k}{2k} 2^{n-k} , \quad n \geqslant 0 . \tag{6.27}$$

We define $A(z)$ to be the ordinary generating function of a_n. We find that

$$\begin{aligned} A(z) &= \sum_{n=0}^{\infty} a_n z^n = \sum_{n=0}^{\infty} z^n \sum_{k=0}^{n} \binom{n+k}{2k} 2^{n-k} \\ &= \sum_{k=0}^{\infty} 2^{-k} \sum_{n=k}^{\infty} 2^n z^n \binom{n+k}{2k} = \sum_{k=0}^{\infty} 2^{-k} (2z)^{-k} \sum_{n=0}^{\infty} \binom{n+k}{2k} (2z)^{n+k} \\ &= \sum_{k=0}^{\infty} 2^{-k} (2z)^{-k} \frac{(2z)^{2k}}{(1-2z)^{2k+1}} = \frac{1}{1-2z} \sum_{k=0}^{\infty} \left(\frac{z}{1-2z} \right)^k \\ &= \frac{1-2z}{(1-4z)(1-z)} = \frac{2}{3(1-4z)} + \frac{1}{3(1-z)} . \end{aligned} \tag{6.28}$$

Therefore we immediately find the explicit form

$$a_n = (2^{2n+1} + 1)/3 \quad \text{for } n \geqslant 0 . \tag{6.29}$$

☒

We next present some additional examples of how generating functions are derived. We start by considering linear recurrences with constant coefficients.

The first step in solving a linear recurrence is to obtain its generating function. Suppose that a sequence $a_0, a_1, a_2, \ldots$ satisfies the recurrence

$$a_n = \sum_{i=1}^{d} c_i a_{n-i}, \quad n \geqslant d \, . \tag{6.30}$$

Then

$$\begin{aligned} f(z) = \sum_{n=0}^{\infty} a_n z^n &= \sum_{n=0}^{d-1} a_n z^n + \sum_{n=d}^{\infty} z^n \sum_{i=1}^{d} c_i a_{n-i} \\ &= \sum_{n=0}^{d-1} a_n z^n + \sum_{i=1}^{d} c_i z^i \sum_{n=d}^{\infty} a_{n-i} z^{n-i} \\ &= \sum_{n=0}^{d-1} a_n z^n + \sum_{i=1}^{d} c_i z^i \left(f(z) - \sum_{n=0}^{d-i-1} a_n z^n \right), \end{aligned} \tag{6.31}$$

and so

$$f(z) = \frac{g(z)}{1 - \sum_{i=1}^{d} c_i z^i}, \tag{6.32}$$

where

$$g(z) = \sum_{n=0}^{d-1} a_n z^n - \sum_{i=1}^{d} c_i z^i \sum_{n=0}^{d-i-1} a_n z^n \tag{6.33}$$

is a polynomial of degree $\leqslant d-1$. Equation (6.32) is the fundamental relation in the study of linear recurrences, and $1 - \sum c_i z^i$ is called the *characteristic polynomial* of the recursion.

Example 6.3 (*Fibonacci numbers*). We let $F_0 = 0$, $F_1 = 1$, $F_n = F_{n-1} + F_{n-2}$ for $n \geqslant 2$, and

$$F(z) = \sum_{n=0}^{\infty} F_n z^n \, .$$

Then by (6.32) and (6.33),

$$F(z) = \frac{z}{1 - z - z^2} \, . \tag{6.34}$$

☒

Often there is no obvious recurrence for the sequence a_n being studied, but there is one involving some other auxiliary function. Usually if one can obtain at least as many recurrences as there are sequences, one can obtain their generating functions by methods similar to those used for a single sequence. The main additional complexity comes from the need to solve a system of linear equations with polynomial coefficients. We illustrate this with the following example.

Example 6.4 (*Sequences with forbidden subwords*). Let $A = a_1a_2 \cdots a_k$ be a binary string of length k. Define $f_A(n)$ to be the number of binary strings of length n that do not contain A as a subword of k adjacent characters. (Subsequences do not count, so that if $A = 1110$, then A is contained in 1101110010, but not in 101101.) We introduce the correlation polynomial $C_A(z)$ of A:

$$C_A(z) = \sum_{j=0}^{k-1} c_A(j)z^j \ , \tag{6.35}$$

where $c_A(0) = 1$ and for $1 \leqslant j \leqslant k-1$,

$$c_A(j) = \begin{cases} 1 & \text{if } a_1a_2 \cdots a_{k-j} = a_{j+1}a_{j+2} \cdots a_k \ , \\ 0 & \text{otherwise} \ . \end{cases} \tag{6.36}$$

As examples, we note that if $A = 1000$, then $C_A(z) = 1$, whereas $C_A(z) = 1 + z + z^2 + z^3$ if $A = 1111$. The generating function

$$F_A(z) = \sum_{n=0}^{\infty} f_A(n)z^n \tag{6.37}$$

then satisfies

$$F_A(z) = \frac{C_A(z)}{z^k + (1-2z)C_A(z)} \ . \tag{6.38}$$

To prove this, define $g_A(n)$ to be the number of binary sequences $b_1b_2 \cdots b_n$ of length n such that $b_1b_2 \cdots b_k = A$, but such that $b_jb_{j+1} \cdots b_{j+k-1} \neq A$ for any j with $2 \leqslant j \leqslant n-k+1$; i.e., sequences that start with A but do not contain it anywhere else. We then have $g_A(n) = 0$ for $n < k$, and $g_A(k) = 1$. We also define

$$G_A(z) = \sum_{n=0}^{\infty} g_A(n)z^n \ . \tag{6.39}$$

We next obtain a relation between $G_A(z)$ and $F_A(z)$ that will enable us to determine both.

If $b_1b_2 \cdots b_n$ is counted by $f_A(n)$, then for x either 0 or 1, the string $xb_1b_2 \cdots b_n$ either does not contain A at all, or if it does contain it, then $A = xb_1b_2 \cdots b_{k-1}$. Therefore for $n \geqslant 0$,

$$2f_A(n) = f_A(n+1) + g_A(n+1) \tag{6.40}$$

and multiplying both sides of eq. (6.40) by z^n and summing on $n \geqslant 0$ yields

$$2F_A(z) = z^{-1}(F_A(z) - 1) + z^{-1}G_A(z) \ . \tag{6.41}$$

We need one more relation, and to obtain it we consider any string $B = b_1 b_2 \cdots b_n$ that does not contain A anywhere inside. If we let C be the concatenation of A and B, so that $C = a_1 a_2 \cdots a_k b_1 b_2 \cdots b_n$, then C starts with A, and may contain other occurrences of A, but only at positions that overlap with the initial A. Therefore we obtain,

$$f_A(n) = \sum_{\substack{j=1 \\ c_A(k-j)=1}}^{k} g_A(n+j) \quad \text{for } n \geqslant 0 \,, \tag{6.42}$$

and this gives the relation

$$F_A(z) = z^{-k} C_A(z) G_A(z) \,. \tag{6.43}$$

Solving the two equations (6.41) and (6.43), we find that $F_A(z)$ satisfies (6.38), while

$$G_A(z) = \frac{z^k}{z^k + (1-2z)C_A(z)} \,. \tag{6.44}$$

The proof above follows that in Guibas and Odlyzko (1981), except that that paper uses generating functions in z^{-1}, so the formulas look different. Applications of the formulas (6.38) and (6.44) will be found later in this chapter, as well as in Guibas and Odlyzko (1981) and Flajolet et al. (1988). Other approaches to string enumeration problems are referenced there as well. Other approaches and applications of string enumerations are given in the references in Guibas and Odlyzko (1981) and in papers such as Arratia et al. (1990b). ☒

The above example can be generalized to provide generating functions that enumerate sequences in which any of a given set of patterns are forbidden (Guibas and Odlyzko 1981).

Whenever one has a finite system of linear recurrences with constant coefficients that involve several sequences, say $a_n^{(i)}$, $1 \leqslant i \leqslant k$, $n \geqslant 0$, one can translate these recurrences into linear equations with polynomial coefficients in the generating functions $A^{(i)}(z) = \sum a_n^{(i)} z^n$ for these sequences. To obtain the $A^{(i)}(z)$, one then needs to solve the resulting system. Such solutions will exist if the matrix of polynomial coefficients is nonsingular over the field of rational functions in z. In particular, one needs at least as many equations (i.e., recurrence relations) as k, the number of sequences, and if there are exactly as many equations as sequences, then the determinant of the matrix of the coefficients has to be a nonzero polynomial.

One interesting observation is that when a system of recurrences involving several sequences is solved by the above method, each of the generating functions $A^{(i)}(z)$ is a rational function in z. What this means is that each of the sequences $a_n^{(i)}$, $1 \leqslant i \leqslant k$, satisfies a linear recurrence with constant coefficients that does not involve any of the other $a_n^{(j)}$ sequences! In principle, therefore, that recurrence could have been found right at the beginning by combinatorial methods. However, usually the degree of the recurrence for an isolated $a_n^{(j)}$ sequence is high, typically about k times as large as the average degree of the k recurrences involving all the $a_n^{(j)}$. Thus the use of several sequences $a_n^{(j)}$ leads to much simpler and combinatorially more appealing relations.

That generating functions can significantly simplify combinatorial problems is shown by the following example. It is taken from Stanley (1978), and is a modification of a result of Klarner (1968) and Pólya (1969). This example also shows a more complicated derivation of explicit generating functions than the simple ones presented so far.

Example 6.5 (*Polyomino enumeration,* Stanley 1978). Let a_n be the number of n-square polyominoes P that are inequivalent under translation, but not necessarily under rotation or reflection, and such that each row of P is an unbroken line of squares. Then $a_1 = 1$, $a_2 = 2$, $a_3 = 6$. We define $a_0 = 0$. It is easily seen that

$$a_n = \sum (m_1 + m_2 - 1)(m_2 + m_3 - 1) \cdots (m_{s-1} + m_s - 1) , \tag{6.45}$$

where the sum is over all ordered partitions $m_1 + \cdots + m_s = n$ of n into positive integers m_i. Let $a_{r,n}$ be the sum of terms in (6.45) with $m_1 = r$, where we set $a_{n,n} = 1$, and $a_{r,n} = 0$ if $r > n$ or $n < 0$. Then

$$a_n = \sum_{r=1}^{\infty} a_{r,n} , \tag{6.46}$$

$$a_{r,n} = \sum_{i=1}^{\infty} (r + i - 1) a_{i,n-r} , \qquad r < n . \tag{6.47}$$

Define

$$A(x,y) = \sum_{n=1}^{\infty} \sum_{r=1}^{\infty} a_{r,n} x^r y^n , \tag{6.48}$$

so that

$$A(1,y) = \sum_{n=1}^{\infty} a_n y^n \tag{6.49}$$

is the generating function of the a_n, which are what we need to estimate.

By (6.47), we find that

$$A(x,y) = \sum_{n=1}^{\infty} x^n y^n + \sum_{n=1}^{\infty} \sum_{r=1}^{\infty} \sum_{i=1}^{\infty} (r + i - 1) a_i(n-r) x^r y^n \tag{6.50}$$

$$= \frac{xy}{1-xy} + \frac{x^2y^2}{(1-xy)^2} A(1,y) + \frac{xy}{1-xy} G(x,y) , \tag{6.51}$$

where

$$G(y) = \sum_{n=1}^{\infty} \sum_{i=1}^{\infty} i a_{i,n} y^n = \left. \frac{\partial}{\partial x} A(x,y) \right|_{x=1} . \tag{6.52}$$

We now set $x = 1$ in (6.50) and obtain an equation involving $A(1, y)$ and $G(y)$, namely

$$A(1,y) = \frac{y}{1-y} + \frac{y^2}{(1-y)^2}A(1,y) + \frac{y}{1-y}G(y) . \tag{6.53}$$

We next differentiate (6.50) with respect to x, and set $x = 1$. This gives us a second equation,

$$G(y) = \frac{y}{(1-y)^2} + \frac{2y^2}{(1-y)^3}A(1,y) + \frac{y}{(1-y)^2}G(y) . \tag{6.54}$$

We now eliminate $G(y)$ from (6.53) and (6.54) to obtain

$$A(1,y) = \frac{y(1-y)^3}{1-5y+7y^2-4y^3} . \tag{6.55}$$

This formula shows that

$$a_{n+3} = a_{n+2} - 7a_{n+1} + 4a_n \quad \text{for } n \geqslant 2 . \tag{6.56}$$

Using the results of section 10 we can easily obtain from (6.55) an asymptotic estimate

$$a_n \sim c\alpha^n \quad \text{as } n \to \infty , \tag{6.57}$$

where c is a certain constant and $\alpha = 3.205569\ldots$ is the inverse of the smallest zero of $1 - 5y + 7y^2 - 4y^3$. ⊠

For other methods and results related to polyomino enumeration, see Privman and Svrakic (1988, 1989).

6.2. *Composition and inversion of power series*

So far we have only discussed simple operations on generating functions, such as multiplication. What happens when we do something more complicated? There are several frequently occurring operations on generating functions whose results can be described explicitly.

Faà di Bruno's formula (Comtet 1974). Suppose that

$$A(z) = \sum_{m=0}^{\infty} a_m \frac{z^m}{m!} , \quad B(z) = \sum_{n=0}^{\infty} b_n \frac{z^n}{n!} , \tag{6.58}$$

are two exponential generating functions with $b_0 = 0$. Then the formal composition $C(z) = A(B(z))$ is well-defined, and

$$C(z) = \sum_{n=0}^{\infty} c_n \frac{z^n}{n!} \tag{6.59}$$

with

$$c_0 = 0, \quad c_n = \sum_{k=1}^{n} a_k B_{n,k}(b_1, b_2, \ldots, b_{n-k+1}) \,, \tag{6.60}$$

where the $B_{n,k}$ are the exponential Bell polynomials defined by

$$\sum_{n,k=0}^{\infty} B_{n,k}(x_1, \ldots, x_{n-k+1}) \frac{t^n u^k}{n!} = \exp\left(u \sum_{m=1}^{\infty} x_m \frac{t^m}{m!} \right) \,, \tag{6.61}$$

with the x_j independent variables.

Faà di Bruno's formula makes it possible to compute successive derivatives of functions such as $\log A(z)$ in terms of the derivatives of $A(z)$. For further examples, see Comtet (1974), Riordan (1958, 1968). Faà di Bruno's formula is derivable in a straightforward way from the multinomial theorem.

Composition of generating functions occurs frequently in combinatorics and analysis of algorithms. When it yields the desired generating function as a composition of several known generating functions, the basic problem is solved, and one can work on the asymptotics of the coefficients using Faà di Bruno's formula or other methods. A more frequent event is that the composition yields a functional equation for the generating function, as in Example 6.1, where the exponential generating function $t(z)$ for labeled rooted trees was shown to satisfy $t(z) = z\exp(t(z))$. General functional equations are hard to deal with. (Many examples will be presented later.) However, there is a class of them for which an old technique, the Lagrange–Bürmann inversion formula, works well. We start by noting that if

$$f(z) = \sum_{n=0}^{\infty} f_n z^n \tag{6.62}$$

is a formal power series with $f_0 = 0$, $f_1 \neq 0$, then there is an inverse formal power series $f^{(-1)}(z)$ such that

$$f(f^{(-1)}(z)) = f^{(-1)}(f(z)) = z \,. \tag{6.63}$$

The coefficients of $f^{(-1)}(z)$ can be expressed explicitly in terms of the coefficients of $f(z)$. More generally, we have the following result.

Lagrange–Bürmann inversion formula. Suppose that $f(z)$ is a formal power series with $[z^0]f(z) = 0$, $[z^1]f(z) \neq 0$, and that $g(z)$ is any formal power series. Then for $n \geqslant 1$,

$$[z^n]\{g(f^{(-1)})(z)\} = n^{-1}[z^{n-1}]\{g'(z)(f(z)/z)^{-n}\} \,. \tag{6.64}$$

In particular, for $g(z) = z$, we have

$$[z^n]f^{(-1)}(z) = n^{-1}[z^{n-1}](f(z)/z)^{-n} \,. \tag{6.65}$$

Example 6.6 (*Rooted labeled trees*). As was shown in Example 6.1, the exponential generating function of rooted labeled trees satisfies $t(z) = z\exp(t(z))$. If we rewrite it as $z = t(z)\exp(-t(z))$, we see that $t(z) = f^{\langle -1 \rangle}(z)$, where $f(z) = z\exp(-z)$. Therefore eq. (6.65) yields

$$\begin{aligned}[z^n]t(z) &= n^{-1}[z^{n-1}]\exp(-nz) \\ &= n^{-1}n^{n-1}/(n-1)! = n^{n-1}/n! \,,\end{aligned} \tag{6.66}$$

which shows that t_n, the number of rooted labeled trees on n nodes, is n^{n-1}. ⊠

Proof of a form of the Lagrange–Bürmann theorem is given in chapter 21. Extensive discussion, proofs, and references are contained in Comtet (1974), Goulden and Jackson (1983), Henrici (1974–86), and Whittaker and Watson (1927). Some additional recent references are Garsia and Joni (1977) and Hofbauer (1979). There exist generalizations of the Lagrange–Bürmann formula to several variables (Goulden and Jackson 1983, Good 1960, Hofbauer 1979).

The Lagrange–Bürmann formula, as stated above, is valid for general formal power series. If $f(z)$ is analytic in a neighborhood of the origin, then so are $f^{\langle -1 \rangle}(z)$ and $g(f^{\langle -1 \rangle})(z)$, provided $g(z)$ is also analytic near 0 and $f'(0) \neq 0$, $f(0) = 0$. Most of the presentations of this inversion formula in the literature assume analyticity. However, that is not a real restriction. To prove (6.65), say, in full generality, it suffices to prove it for any n. Given n, if we let

$$F(z) = \sum_{k=0}^{n} f_k z^k \,, \quad G(z) = \sum_{k=0}^{n} g_k z^k \,,$$

then we see that

$$[z^n]\{g(f^{\langle -1 \rangle})(z)\} = [z^n]G(F^{\langle -1 \rangle})(z) \,, \tag{6.67}$$

and $F(z)$ and $G(z)$ are analytic, so the formula (6.65) can be applied. Thus combinatorial proofs of the Lagrange–Bürmann formula do not offer greater generality than analytic ones.

While the analytic vs. combinatorial distinction in the proofs of the Lagrange–Bürmann formula does not matter, it is possible to use analyticity of the functions $f(z)$ and $g(z)$ to obtain useful information. Example 6.6 above was atypical in that a simple explicit formula was derived. Often the quantity on the right-hand side of (6.64) is not explicit enough to make clear its asymptotic behavior. When that happens, and $g(z)$ and $f(z)$ are analytic, one can use the contour integral representation

$$[z^{n-1}]\{g'(z)(f(z)/z)^{-n}\} = \frac{1}{2\pi i}\int_{\Gamma} g'(z)f(z)^{-n}\,dz \,, \tag{6.68}$$

where Γ is a positively oriented simple closed contour enclosing the origin that lies inside the region of analyticity of both $g(z)$ and $f(z)$. This representation, which is discussed in section 10, can often be used to obtain asymptotic information about coefficients $[z^n]g(f^{\langle -1\rangle})(z)$ (cf. Mallows et al. 1975).

The Lagrange–Bürmann formula can provide numerical approximations to roots of equations and even convergent infinite-series representations for such roots. An important case is the trinomial equation $y = z(1 + y^r)$, and there are many others.

Example 6.7 (*Dominant zero for forbidden-subword generating functions*). The generating functions $F_A(z)$ and $G_A(z)$ of Example 6.4 both have denominators

$$h(z) = z^k + (1 - 2z)C(z) , \tag{6.69}$$

where $C(z)$ is a polynomial of degree $\leqslant k$, with coefficients 0 and 1, and with $C(0) = 1$. It will be shown later that $h(z)$ has only one zero ρ of small absolute value, and that this zero is the dominant influence on the asymptotic behavior of the coefficients of $F_A(z)$ and $G_A(z)$. Right now we obtain accurate estimates for ρ.

For simplicity, we will consider only large k. Since $C(z)$ has nonnegative coefficients and $C(0) = 1$, $h(3/4) \leqslant (3/4)^k - 1/2 < 0$ for $k \geqslant 3$. On the other hand, $h(1/2) = 2^{-k}$. Therefore $h(z)$ has a real zero ρ with $1/2 < \rho < 3/4$. As $k \to \infty$, $\rho \to 1/2$, since

$$\rho^k = (2\rho - 1)C(\rho) , \tag{6.70}$$

and $\rho^k \to 0$ as $k \to \infty$ for $1/2 < \rho < 3/4$, while $2\rho - 1$ and $C(\rho)$ are bounded. We can deduce from (6.69) that

$$2\rho - 1 \sim 2^{-k}C(1/2)^{-1} \quad \text{as } k \to \infty , \tag{6.71}$$

uniformly for all polynomials $C(z)$ of the prescribed type. By applying the bootstrapping technique (see section 5.4) we can find even better approximations. By (6.71),

$$C(\rho) = C(1/2) + O(|\rho - 1/2|) = C(1/2) + O(2^{-k}) , \tag{6.72}$$

$$\rho^k = 2^{-k}(1 + O(2^{-k}))^k = 2^{-k}(1 + O(k2^{-k})) , \tag{6.73}$$

so (6.70) now yields

$$\rho = 1/2 + 2^{-k-1}C(1/2)^{-1} + O(k2^{-2k}) . \tag{6.74}$$

Even better approximations can be obtained by repeating the process using (6.74). At the next stage we would apply the expansion

$$\begin{aligned} C(\rho) &= C(1/2) + (\rho - 1/2)C'(1/2) + O((\rho - 1/2)^2) \\ &= C(1/2) + 2^{-k-1}C'(1/2) + O(k2^{-2k}) \end{aligned} \tag{6.75}$$

and a similar one for ρ^k.

A more systematic way to obtain a rapidly convergent series for ρ is to use the inversion formula. If we set $u = \rho - 1/2$, then (6.70) can be rewritten as $w(u) = 1$, where

$$w(u) = 2uC(1/2+u)(1/2+u)^{-k} = \sum_{j=1}^{\infty} a_j u^j \ , \tag{6.76}$$

with

$$a_1 = 2^{k+1}C(1/2) \neq 0 \ . \tag{6.77}$$

Hence $u = w^{\langle -1 \rangle}(1)$, and the Lagrange–Bürmann inversion formula (6.65) yields the coefficients of $w^{\langle -1 \rangle}(z)$. In particular, we find that

$$\begin{aligned} \rho = 1/2 + u \approx 1/2 + 2^{-k-1}C(1/2)^{-1} + k2^{-2k-1}C(1/2)^{-2} \\ -2^{-2k-2}C'(1/2)C(1/2)^{-3} + \cdots \end{aligned} \tag{6.78}$$

as a Poincaré asymptotic series. With additional work one can show that the series (6.78) converges, and that

$$\begin{aligned} \rho = 1/2 + 2^{-k-1}C(1/2)^{-1} + k2^{-2k-1}C(1/2)^{-2} \\ -2^{-2k-2}C'(1/2)C(1/2)^{-3} + O(k^2 2^{-3k}) \ , \end{aligned} \tag{6.79}$$

for example. The same estimate can be obtained by the bootstrapping technique. ⊠

6.3. *Differentiably finite power series*

Homogeneous recurrences with constant coefficients are the nicest large set of sequences one can imagine, with rational generating functions, and well-understood asymptotic behavior. The next class in complexity consists of the polynomially recursive or, *P-recursive sequences*, $a_0, a_1, \ldots$, which satisfy recurrences of the form

$$p_d(n)a_{n+d} + p_{d-1}(n)a_{n+d-1} + \cdots + p_0(n)a_n = 0, \quad n \geqslant 0 \ , \tag{6.80}$$

where d is fixed and $p_0(n), \ldots, p_d(n)$ are polynomials in n. Such sequences are common in combinatorics, with $a_n = n!$ a simple example. Normally P-recursive sequences do not have explicit forms for their generating functions. In this section we briefly summarize some of their main properties. Asymptotic properties of P-recursive sequences will be discussed in section 9.2. The main references for the results quoted here are Lipshitz (1989) and Stanley (1980).

A formal power series

$$f(z) = \sum_{k=0}^{\infty} a_k z^k \tag{6.81}$$

is called differentiably finite, or *D-finite*, if the derivatives $f^{(n)}(z) = \frac{\mathrm{d}^n f(z)}{\mathrm{d}z^n}$, $n \geqslant 0$, span a finite-dimensional vector space over the field of rational functions with complex coefficients. The following three conditions are equivalent for a formal power series $f(z)$:

(i) $f(z)$ is D-finite.

(ii) There exist finitely many polynomials $q_0(z), \ldots, q_k(z)$ and a polynomial $q(z)$, not all 0, such that

$$q_k(z)f^{(k)}(z) + \cdots + q_0(z)f(z) = q(z) . \tag{6.82}$$

(iii) There exist finitely many polynomials $p_0(z), \ldots, p_m(z)$, not all 0, such that

$$p_m(z)f^{(m)}(z) + \cdots + p_0(z)f(z) = 0 . \tag{6.83}$$

The most important result for combinatorial enumeration is that a sequence $a_0, a_1, \ldots,$ is P-recursive if and only if its ordinary generating function $f(z)$, defined by (6.81), is D-finite. This makes it possible to apply results that are more easily proved for D-finite power series.

If $f(z)$ is D-finite, then so is the power series obtained by changing a finite number of the coefficients of $f(z)$. If $f(z)$ is algebraic (i.e., there exist polynomials $q_0(z), \ldots, q_d(z)$, not all 0, such that $q_d(z)f(z)^d + \cdots + q_0(z)f(z) + q_0(z) = 0$), then $f(z)$ is D-finite. The product of two D-finite power series is also D-finite, as is any linear combination with polynomial coefficients. Finally, the Hadamard product of two D-finite series is D-finite. The proofs rely on elementary linear-algebra constructions. An important feature of the theory is that identity between D-finite series is decidable.

The concept of a D-finite power series can be extended to several variables (Lipshitz 1989, Zeilberger 1990), and there are generalizations of P-recursiveness (Lipshitz 1989, Zeilberger 1990). (See also Gessel 1990.) Zeilberger (1990) has used the word *holonomic* to describe corresponding sequences and generating functions.

When we investigate a sequence $\{a_n\}$, sometimes the combinatorial context yields only relations for a more complicated object with several indices. While we might like to obtain the generating function $f(z) = \sum a_n z^n$, we might instead find a formula for a generating function

$$F(z_1, z_2, \ldots, z_k) = \sum_{n_1,\ldots,n_k} b_{n_1}, \ldots, {}_{n_k} z_1^{n_1}, \ldots, z_k^{n_k} , \tag{6.84}$$

where $a_n = b_{n,n,\ldots,n}$, say. When this happens, we say that $f(z)$ is a *diagonal* of $F(z_1, \ldots, z_k)$. [There are more general definitions of diagonals in Denef and Lipshitz (1987), Lipshitz (1988, 1989), and Lipshitz and van der Poorten (1990), which are recent references for this topic.] Diagonals of D-finite power series in any number of variables are D-finite. Diagonals of two-variable rational functions are algebraic, but there are three-variable rational functions whose diagonals are not algebraic (Furstenberg 1967).

6.4. *Unimodality and log-concavity*

A finite sequence $a_0, a_1, \ldots, a_n$ of real numbers is called *unimodal* if for some index k, $a_0 \leqslant a_1 \leqslant \cdots \leqslant a_k$ and $a_k \geqslant a_{k+1} \geqslant \cdots \geqslant a_n$. A sequence $a_0, \ldots, a_n$ of nonnegative elements is called *log-concave* (short for logarithmically concave) if $a_j^2 \geqslant a_{j-1}a_{j+1}$ holds for $1 \leqslant j \leqslant n-1$. Unimodal and log-concave sequences occur frequently in combinatorics and are objects of intensive study. We present a brief review of some of their properties because asymptotic methods are often used to prove unimodality and log-concavity. Furthermore, knowledge that a sequence is log-concave or unimodal is often helpful in obtaining asymptotic information. For example, some methods provide only asymptotic estimates for summatory functions of sequences, and unimodality helps in obtaining from those estimates bounds on individual coefficients. This approach will be presented in section 13, in the discussion of central and local limit theorems.

The basic references for unimodality and log-concavity are Karlin (1968) and Stanley (1989). For recent results, see also Brenti (1989) and the references given there. All the results listed below can be found in those sources and the references they list.

In the rest of this subsection we will consider only sequences of nonnegative elements. A sequence $a_0, \ldots, a_n$ will be said to *have no internal zeros* if there is no triple of integers $0 \leqslant i < j < k \leqslant n$ such that $a_j = 0$, $a_i a_k \neq 0$. It is easy to see that a log-concave sequence with no internal zeros is unimodal, but there are sequences of positive elements that are unimodal but not concave. The convolution of two unimodal sequences does not have to be unimodal. However, it is unimodal if each of the two unimodal sequences is also symmetric. Convolution of two log-concave sequences is log-concave. The convolution of a log-concave and a unimodal sequence is unimodal. A log-concave sequence is even characterized by the property that its convolution with any unimodal sequence is unimodal. This last property is related to the variation-diminishing character of log-concave sequences (see Karlin 1968), which we will not discuss at greater length here except to note that there are more restrictive sets of sequences (the Pólya frequency classes, see Brenti 1989, Karlin 1968) which have stronger convolution properties.

The binomial coefficients $\binom{n}{k}$, $0 \leqslant k \leqslant n$, are log-concave, and therefore unimodal. The q-binomial coefficients $\left[{n \atop k}\right]_q$ are log-concave for any $q \geqslant 1$. On the other hand, if we write a single coefficient $\left[{n \atop k}\right]_q$ for fixed n and k as a polynomial in q, the sequence of coefficients is unimodal, but does not have to be log-concave.

The most frequently used method for showing that a sequence $a_0, \ldots, a_n$ is log-concave is to show that all the zeros of the polynomial

$$A(z) = \sum_{k=0}^{n} a_k z^k \tag{6.85}$$

are real and $\leqslant 0$. In that case not only are the a_k log-concave, but so are $a_k \binom{n}{k}^{-1}$. Absolute values of the Stirling numbers of both kinds were first shown to be log-concave by this method (Harper 1967). There are many unsolved conjectures about log-concavity of combinatorial sequences, such as the Read–Hoggar conjecture that coefficients of chromatic polynomials are log-concave (cf. Brenti et al. 1994).

A variety of combinatorial, algebraic, and geometric methods have been used to prove unimodality of sequences, and we refer the reader to Stanley (1989) for a comprehensive and insightful survey. In section 12.3 we will discuss briefly some proofs of unimodality and log-concavity that use asymptotic methods. The basic philosophy is that since the Gaussian distribution is log-concave and unimodal (when we extend the definition of these concepts to continuous distributions), these properties should also hold for sequences that by the central limit theorem or its variants are asymptotic to the Gaussian. Therefore one can expect high-order convolutions of sequences to be log-concave at least in their central region, and there are theorems that prove this under certain conditions.

6.5. *Moments and distributions*

The second-moment method is a frequently used technique in probabilistic arguments, as is shown in chapter 33 and Bollobás (1985), Erdős and Spencer (1974), and Spencer (1987). It is based on *Chebyshev's inequality*, which says that if X is a real-valued random variable with finite second moment $E(X^2)$, then

$$\mathrm{Prob}\,(|X - E(X)| \geqslant \alpha|E(X)|) \leqslant \frac{E(X^2) - E(X)^2}{\alpha^2 E(X)^2}\,. \tag{6.86}$$

An easy corollary of inequality (6.86) that is often used is

$$\mathrm{Prob}\,(X = 0) \leqslant \frac{E(X^2) - E(X)^2}{E(X)^2}\,. \tag{6.87}$$

[There is a slightly stronger version of the inequality (6.87), in which $E(X)^2$ in the denominator is replaced by $E(X^2)$.] The inequalities (6.86) and (6.87) are usually applied for $X = Y_1 + \cdots + Y_n$, where the Y_j are other random variables. The helpful feature of the inequalities is that they require only knowledge of the pairwise dependencies among the Y_j, which is easier to study than the full joint distribution of the Y_j. For other bounds on distributions that can be obtained from partial information about moments, see Shohat and Tamarkin (1943).

The reason moment bounds are mentioned at all in this chapter is that asymptotic methods are often used to derive them. Generating functions are a common and convenient method for doing this.

Example 6.8 (*Waiting times for subwords*). In a continuation and application of Example 6.4, let A be a binary string of length k. How many tosses of a fair coin (with sides labeled 0 and 1) are needed on average before A appears as a block of k consecutive outcomes? By a general observation of probability theory, this is just the sum over $n \geqslant 0$ of the probability that A does not appear in the first n coin tosses, and thus equals

$$\sum_{n\ 0}^{\infty} f_A(n)2^{-n} = F_A(1/2) = 2^k C_A(1/2)\,, \tag{6.88}$$

where the last equality follows from eq. (6.38). Another, more general, way to derive this is to use $G_A(z)$. Note that $g_A(n)2^{-n}$ is the probability that A appears in the first n coin tosses, but not in the first $n-1$. Hence the rth moment of the time until A appears is

$$\sum_{n=0}^{\infty} n^r g_A(n) 2^{-n} = \left(z\frac{\mathrm{d}}{\mathrm{d}z}\right)^r G_A(z)\bigg|_{z=1/2} . \tag{6.89}$$

If we take $r=1$, we again obtain the expected waiting time given by (6.88). When we take $r=2$, we find that the second moment of the time until the appearance of A is

$$\sum_{n=0}^{\infty} n^2 g_A(n) 2^{-n} = 2^{2k+1} C_A(1/2)^2 - (2k-1)2^k C_A(1/2) + 2^k C_A'(1/2) , \tag{6.90}$$

and therefore the variance is

$$\begin{aligned} &2^{2k} C_A(1/2)^2 - (2k-1)2^k C_A(1/2) + 2^k C_A'(1/2) \\ &\quad = 2^{2k} C_A(1/2)^2 + \mathrm{O}(k2^k) , \end{aligned} \tag{6.91}$$

since $1 \leqslant C_A(1/2) \leqslant 2$. Higher moments can be used to obtain more detailed information. A better approach is to use the method of Example 9.3, which gives precise estimates for the tails as well as the mean of the distribution. ⊠

Information about moments of distribution functions can often be used to obtain the limiting distribution. If $F_n(x)$ is a sequence of distribution functions such that for every integer $k \geqslant 0$, the kth moment

$$\mu_n(k) = \int x^k \,\mathrm{d}F_n(x) \tag{6.92}$$

converges to $\mu(k)$ as $n \to \infty$, then there is a limiting measure with distribution function $F(x)$ whose kth moment is $\mu(k)$. If the moments $\mu(k)$ do not grow too rapidly, then they determine the distribution function $F(x)$ uniquely, and the $F_n(x)$ converge to $F(x)$ (in the weak star sense, Billingsley 1979). A sufficient condition for the $\mu(k)$ to determine $F(x)$ uniquely is that the generating function

$$U(x) = \sum_{k=0}^{\infty} \frac{\mu(2k)x^k}{(2k)!} \tag{6.93}$$

should converge for some $x > 0$. In particular, the standard normal distribution with

$$F(x) = (2\pi)^{-1/2} \int_{\infty}^{x} \exp(-u^2/2)\,\mathrm{d}u \tag{6.94}$$

has $\mu(2k) = 1 \cdot 3 \cdot 5 \cdot 7 \cdot \cdots \cdot (2k-1)$ (and $\mu(2k+1) = 0$), so it is determined uniquely by its moments. On the other hand, there are some frequently encountered distributions, such as the log-normal one, which do not have this property.

7. Formal power series

This section discusses generating functions $f(z)$ that might not converge in any interval around the origin. Sequences that grow rapidly are common in combinatorics, with $a_n = n!$ the most obvious example for which

$$f(z) = \sum_{n=0}^{\infty} a_n z^n \tag{7.1}$$

does not converge for any $z \neq 0$. The usual way to deal with the problem of a rapidly growing sequence a_n is to study the generating function of a_n/b_n, where b_n is some sequence with known asymptotic behavior. When $b_n = n!$, the ordinary generating function of a_n/b_n is then the exponential generating function of a_n. For derangements [eqs. (1.1) and (6.7)] this works well, as the exponential generating function of d_n converges in $|z| < 1$ and has a nice form. Unfortunately, while we can always find a sequence b_n that will make the ordinary generating function $f(z)$ of a_n/b_n converge (even for all z), usually we cannot do it in a way that will yield any useful information about $f(z)$. The combinatorial structure of a problem almost always severely restricts what forms of generating function can be used to take advantage of the special properties of the problem. This difficulty is common, for example, in enumeration of labeled graphs. In such cases one often resorts to formal power series that do not converge in any neighborhood of the origin. For example, if $c(n,k)$ is the number of connected labeled graphs on n vertices with k edges, then it is well known (cf. Stanley 1978) that

$$\sum_{n=0}^{\infty}\sum_{k=0}^{\infty} c(n,k)\frac{x^k y^n}{n!} = \log\left(\sum_{m=0}^{\infty} \frac{(1+x)^{\binom{m}{2}} y^m}{m!}\right) . \tag{7.2}$$

While the series inside the log in (7.2) does converge for $-2 \leqslant x \leqslant 0$, and any y, it diverges for any $x > 0$ as long as $y \neq 0$, and so this is a relation of formal power series.

There are few methods for dealing with asymptotics of formal power series, at least when compared to the wealth of techniques available for studying analytic generating functions. Fortunately, combinatorial enumeration problems that do require the use of formal power series often involve rapidly growing sequences of positive terms, for which some simple techniques apply. We start with an easy general result that is applicable both to convergent and purely formal power series.

Theorem 7.1 (Bender 1974). *Suppose that $a(z) = \sum a_n z^n$ and $b(z) = \sum b_n z^n$ are power series with radii of convergence $\alpha > \beta \geqslant 0$, respectively. Suppose that $b_{n-1}/b_n \to \beta$ as $n \to \infty$. If $a(\beta) \neq 0$, and $\sum c_n z^n = a(z)b(z)$, then*

$$c_n \sim a(\beta) b_n \quad \textit{as } n \to \infty . \tag{7.3}$$

The proof of Theorem 7.1, which can be found in Bender (1974), is simple. The condition $\alpha > \beta$ is important, and cannot be replaced by $\alpha = \beta$. We can have $\beta = 0$, and that is indeed the only possibility if the series for $b(z)$ does not converge in a neighborhood of $z = 0$.

Example 7.2 (*Double set coverings,* Bender 1974, Comtet 1968). Let v_n be the number of choices of subsets $S_1, \ldots, S_r$ of an n-element set T such that each $t \in T$ is in exactly two of the S_i. There is no restriction on r, the number of subsets, and some of the S_i can be repeated. Let c_n be the corresponding number when the S_i are required to be distinct. We let $C(z) = \sum c_n z^n/n!$, $V(z) = \sum v_n z^n/n!$ be the exponential generating functions. Then it can be shown that

$$C(z) = \exp(-1 - (e^z - 1)/2)A(z)\ , \tag{7.4}$$

$$V(z) = \exp(-1 + (e^z - 1)/2)A(z)\ , \tag{7.5}$$

where

$$A(z) = \sum_{k=0}^{\infty} \exp(k(k-1)z/2)/k!\ . \tag{7.6}$$

We see immediately that $A(z)$ does not converge in any neighborhood of the origin. We have

$$a_n = [z^n]A(z) = 2^{-n} \sum_{k=2}^{\infty} \frac{k^n(k-1)^n}{k!}\ . \tag{7.7}$$

By considering the ratio of consecutive terms in the sum in (7.7), we find that the largest term occurs for $k = k_0$ with $k_0 \log k_0 \sim 2n$, and by the methods of section 5.1 we find that

$$a_n \sim \frac{\pi^{1/2} k_0^n (k_0 - 1)^n}{n^{1/2} 2^n (k_0 - 1)!} \quad \text{as } n \to \infty\ . \tag{7.8}$$

Therefore $a_{n-1}/a_n \to 0$ as $n \to \infty$, and Theorem 7.1 tells us that

$$c_n \sim v_n \sim e^{-1} n! a_n \quad \text{as } n \to \infty\ . \tag{7.9}$$

☒

Usually formal power series occur in more complicated relations than those covered by Theorem 7.1. For example, if f_n is the number of connected graphs on n labeled vertices which have some property, and F_n is the number of graphs on n labeled vertices each of whose connected components has that property, then (cf. Wright 1970)

$$1 + \sum_{n=1}^{\infty} F_n \frac{x^n}{n!} = \exp\left(\sum_{n=1}^{\infty} f_n \frac{x^n}{n!} \right)\ . \tag{7.10}$$

Theorem 7.3 (Bender 1975). *Suppose that*

$$\begin{aligned} a(x) &= \textstyle\sum_{n=1}^{\infty} a_n x^n\ , \qquad F(x,y) = \sum_{h,k \geqslant 0} f_{hk} x^h y^k\ , \\ b(x) &= \textstyle\sum_{n=0}^{\infty} b_n x^n = F(x, a(x))\ , \qquad D(x) = F_y(x, a(x))\ , \end{aligned} \tag{7.11}$$

where $F_y(x,y)$ is the partial derivative of $F(x,y)$ with respect to y. Assume that $a_n \neq 0$ and

$$\text{(i)} \quad a_{n-1} = o(a_n) \quad \textit{as } n \to \infty\ , \tag{7.12}$$

$$\text{(ii)} \quad \sum_{k=r}^{n-r} |a_k a_{n-k}| = O(a_{n-r}) \quad \textit{for some } r > 0\ , \tag{7.13}$$

(iii) *for every $\delta > 0$ there are $M(\delta)$ and $K(\delta)$ such that for $n \geqslant M(\delta)$ and $h + k > r + 1$,*

$$|f_{hk} a_{n-h-k+1}| \leqslant K(\delta)\delta^{h+k}|a_{n-r}|\ . \tag{7.14}$$

Then

$$b_n = \sum_{k=0}^{r-1} d_k a_{n-k} + O(a_{n-r})\ . \tag{7.15}$$

Condition (iii) of Theorem 7.3 is often hard to verify. Theorem 2 of Bender (1975) shows that this condition holds under certain simpler hypotheses. It follows from that result that (iii) is valid if $F(x,y)$ is analytic in x and y in a neighborhood of $(0,0)$. Hence, if $F(x,y) = \exp(y)$ or $F(x,y) = 1 + y$, then Theorem 7.3 becomes easy to apply. One can also deduce from Theorem 2 of Bender (1975) that Theorem 7.3 applies when (i) and (ii) hold, $b_0 = 0$, $b_n \geqslant 0$, and

$$1 + a(z) = \exp\left(\sum_{k=1}^{\infty} b(z^k)/k\right)\ , \tag{7.16}$$

another relation that is common in graph enumeration (cf. Example 15.1). There are also some results weaker than Theorem 7.3 that are easier to apply (Wright 1967).

Example 7.4 (*Indecomposable permutations*, Comtet 1974). For every permutation σ of $\{1,\ldots,n)$, let $\{1,\ldots,n\} = \cup I_h$, where the I_h are the smallest intervals such that $\sigma(I_h) = I_h$ for all h. For example, $\sigma = (134)(2)(56)$ corresponds to $I_1 = \{1,2,3,4\}$, $I_2 = \{5,6\}$, and the identity permutation has n components. A permutation is said to be indecomposable if it has one component. For example, if σ has the 2-cycle $(1n)$, it is indecomposable. Let c_n be the number of indecomposable permutations of $\{1,\ldots,n\}$. Then (Comtet 1974):

$$\sum_{n=1}^{\infty} c_n z^n = 1 - \frac{1}{1 + \sum_{n=1}^{\infty} n! z^n}\ . \tag{7.17}$$

We apply Theorem 7.3 with $a_n = n!$ for $n \geqslant 1$ and $F(x,y) = 1 - (1+y)^{-1}$. We easily obtain

$$c_n \sim n! \quad \text{as } n \to \infty\ , \tag{7.18}$$

so that almost all permutations are indecomposable. ☒

Some further useful expansions for functional inverses and computations of formal power series have been obtained by Bender and Richmond (1984).

8. Elementary estimates for convergent generating functions

The word "elementary" in the title of this section is a technical term that means the proofs do not use complex variables. It does not necessarily imply that the proofs are simple. While some, such as those of section 8.4, are easy, others are more complicated. The main advantage of elementary methods is that they are much easier to use, and since they impose much weaker requirements on the generating functions, they are more widely applicable. Usually they only impose conditions on the generating function $f(z)$ for $z \in \mathbb{R}^+$.

The main disadvantage of elementary methods is that the estimates they give tend to be much weaker than those derived using analytic function approaches. It is easy to explain why that is so by considering the two generating functions

$$f_1(z) = \sum_{n=0}^{\infty} z^n = (1-z)^{-1} \tag{8.1}$$

and

$$f_2(z) = 3/2 + \sum_{n=1}^{\infty} 2z^{2n} = 3/2 + 2z^2(1-z^2)^{-1} \,. \tag{8.2}$$

Both series converge for $|z| < 1$ and diverge for $|z| > 1$, and both blow up as $z \to 1$. However,

$$f_1(z) - f_2(z) = -\frac{1-z}{2(1+z)} \to 0 \quad \text{as } z \to 1 \,. \tag{8.3}$$

Thus these two functions behave almost identically near $z = 1$. Since $f_1(z)$ and $f_2(z)$ are both $\sim (1-z)^{-1}$ as $z \to 1^-$, $z \in \mathbb{R}^+$, and their difference is $O(|z-1|)$ for $z \in \mathbb{R}^+$, it would require exceptionally delicate methods to detect the differences in the coefficients of the $f_j(z)$ just from their behavior for $z \in \mathbb{R}^+$. There is a substantial difference in the behavior of $f_1(z)$ and $f_2(z)$ for real z if we let $z \to -1$, so our argument does not completely exclude the possibility of obtaining detailed information about the coefficients of these functions using methods of real variables only. However, if we consider the function

$$f_3(z) = 2 + \sum_{n=1}^{\infty} 3z^{3n} = 2 + 3z^3(1-z^3)^{-1} \,, \tag{8.4}$$

then $f_1(z)$ and $f_3(z)$ are both $\sim (1-z)^{-1}$ as $z \to 1^-$, $z \in \mathbb{R}^+$, yet now

$$|f_1(z) - f_3(z)| = O(|z-1|) \quad \text{for all } z \in \mathbb{R} \,.$$

This difference is comparable to what would be obtained by modifying a single coefficient of one generating function. To determine how such slight changes in the behavior of the generating functions affect the behavior of the coefficients we would need to know much more about the functions if we were to use real-variable methods. On the other hand, analytic methods, discussed in section 10 and later, are good at dealing with such problems. They require less precise knowledge of the behavior of a function on the real line. Instead, they impose weak conditions on the function in a wider domain, namely that of the complex numbers.

For reasons discussed above, elementary methods cannot be expected to produce precise estimates of individual coefficients. They often do produce good estimates of summatory functions of the coefficients, though. In the examples above, we note that

$$\sum_{n=1}^{N} [z^n] f_j(z) \sim N \quad \text{as } N \to \infty \tag{8.5}$$

for $1 \leqslant j \leqslant 3$. This holds because the $f_j(z)$ have the same behavior as $z \to 1^-$, and is part of a more general phenomenon. Good knowledge of the behavior of the generating function on the real axis combined with weak restrictions on the coefficients often leads to estimates for the summatory function of the coefficients.

There are cases where elementary methods give precise bounds for individual coefficients. Typically when we wish to estimate f_n, with ordinary generating function $f(z) = \sum f_n z^n$ that converges for $|z| < 1$ but not for $|z| > 1$, we apply the methods of this section to

$$g_n = f_n - f_{n-1} \quad \text{for } n \geqslant 1, \quad g_0 = f_0 \tag{8.6}$$

with generating function

$$g(z) = \sum_{n=0}^{\infty} g_n z^n = (1 - z) f(z) \,. \tag{8.7}$$

Then

$$\sum_{k=0}^{n} g_k = f_n \,, \tag{8.8}$$

and so estimates of the summatory function of the g_k yield estimates for f_n. The difficulty with this approach is that now $g(z)$ and not $f(z)$ has to satisfy the hypotheses of the theorems, which requires more knowledge of the f_n. For example, most of the Tauberian theorems apply only to power series with nonnegative coefficients. Hence to use the differencing trick above to obtain estimates for f_n we need to know that $f_{n-1} \leqslant f_n$ for all n. In some cases (such as that of $f_n = p_n$, the number of ordinary partitions of n) this is easily seen to hold through combinatorial arguments. In other situations where one might like to apply elementary methods, though, $f_{n-1} \leqslant f_n$ is either false or else is hard to prove. When that happens, other methods are required to estimate f_n.

8.1. Simple upper and lower bounds

A trivial upper bound method turns out to be widely applicable in asymptotic enumeration, and is surprisingly powerful. It relies on nothing more than the nonnegativity of the coefficients of a generating function.

Lemma 8.1. *Suppose that $f(z)$ is analytic in $|z| < R$, and that $[z^n]f(z) \geqslant 0$ for all $n \geqslant 0$. Then for any x, $0 < x < R$, and any $n \geqslant 0$,*

$$[z^n]f(z) \leqslant x^{-n}f(x) . \tag{8.9}$$

Example 8.2 (*Lower bound for factorials*). Let $f(z) = \exp(z)$. Then Lemma 8.1 yields

$$\frac{1}{n!} = [z^n]\,\mathrm{e}^z \leqslant x^{-n}\,\mathrm{e}^x \tag{8.10}$$

for every $x > 0$. The logarithm of $x^{-n}\mathrm{e}^x$ is $x - n\log x$, and differentiating and setting it equal to 0 shows that the minimum value is attained at $x = n$. Therefore

$$\frac{1}{n!} = [z^n]\,\mathrm{e}^z \leqslant n^{-n}\,\mathrm{e}^n , \tag{8.11}$$

and so $n! \geqslant n^n\,\mathrm{e}^{-n}$. This lower bound holds uniformly for all n, and is off only by an asymptotic factor of $(2\pi n)^{1/2}$ from Stirling's formula (4.1). ☒

Suppose that $f(z) = \sum f_n z^n$. Lemma 8.1 is proved by noting that for $0 < x < R$, the nth term, $f_n x^n$, in the power series expansion of $f(x)$, is $\leqslant f(x)$. As we will see in section 10, it is often possible to derive a similar bound on the coefficients f_n even without assuming that they are nonnegative. However, the proof of Lemma 8.1 shows something more, namely that

$$f_0 x^{-n} + f_1 x^{-n+1} + \cdots + f_{n-1}x^{-1} + f_n \leqslant x^{-n}f(x) \tag{8.12}$$

for $0 < x < R$. When $x \leqslant 1$, this yields an upper bound for the summatory function of the coefficients. Because (8.12) holds, we see that the bound of Lemma 8.1 cannot be sharp in general. What is remarkable is that the estimates obtainable from that lemma are often not far from best possible.

Example 8.3 (*Upper bound for the partition function*). Let $p(n)$ denote the partition function. It has the ordinary generating function

$$f(z) = \sum_{n=0}^{\infty} p(n)z^n = \prod_{k=1}^{\infty}(1 - z^k)^{-1} . \tag{8.13}$$

Let $g(s) = \log f(\mathrm{e}^{-s})$, and consider $s > 0$, $s \to 0$. There are extremely accurate estimates of $g(s)$. It is known (Andrews 1976, Ayoub 1963), for example, that

$$g(s) = \pi^2/(6s) + (\log s)/2 - (\log 2\pi)/2 - s/24 + \mathrm{O}(\exp(-4\pi^2/s)) . \tag{8.14}$$

If we use (8.14), we find that $x^{-n}f(x)$ is minimized at $x = \exp(-s)$ with

$$s = \pi/(6n)^{1/2} - 1/(4n) + \mathrm{O}(n^{-3/2}) , \tag{8.15}$$

which yields

$$p(1)+p(2)+\cdots+p(n) \leqslant 2^{-3/4}\mathrm{e}^{-1/4}n^{-1/4}(1+\mathrm{o}(1))\exp(2\pi 6^{-1/2}n^{1/2})\,. \tag{8.16}$$

Comparing this to the asymptotic formula for the sum that is obtainable from (1.6) (see Example 5.2), we see that the bound of (8.16) is too high by a factor of $n^{1/4}$. If we use (8.16) to bound $p(n)$ alone, we obtain a bound that is too large by a factor of $n^{3/4}$.

The application of Lemma 8.1 outlined above depended on the expansion (8.14), which is complicated to derive, involving modular transformation properties of $p(n)$ that are beyond the scope of this survey. [See Andrews (1976) or Ayoub (1963) for derivations.] Weaker estimates that are still useful are much easier to derive. We obtain one such bound here, since the arguments illustrate some of the methods from the preceding sections.

Consider

$$g(s)=\sum_{k=1}^{\infty}-\log(1-\mathrm{e}^{-ks})\,. \tag{8.17}$$

If we replace the sum by the integral

$$I(s)=\int_1^{\infty}-\log(1-\mathrm{e}^{-us})\,\mathrm{d}u\,, \tag{8.18}$$

we find on expanding the logarithm that

$$I(s)=\int_1^{\infty}\left(\sum_{m=1}^{\infty}m^{-1}\mathrm{e}^{-mus}\right)\mathrm{d}u=s^{-1}\sum_{m=1}^{\infty}m^{-2}\mathrm{e}^{-ms}\,, \tag{8.19}$$

since the interchange of summation and integration is easy to justify, as all the terms are positive. Therefore as $s \to 0^+$,

$$sI(s) \to \sum_{m=1}^{\infty}m^{-2}=\pi^2/6\,, \tag{8.20}$$

so that $I(s) \sim \pi^2/(6s)$ as $s \to 0^+$. It remains to show that I is indeed a good approximation to $g(s)$. This follows easily from the bound (5.32), since it shows that

$$g(s)=I(s)+\mathrm{O}\left(\int_1^{\infty}\frac{s\,\mathrm{e}^{-vs}}{1-\mathrm{e}^{-vs}}\,\mathrm{d}v\right)\,. \tag{8.21}$$

We could estimate the integral in (8.21) carefully, but we only need rough upper bounds for it, so we write it as

$$\int_1^{\infty}\frac{s\,\mathrm{e}^{-vs}}{1-\mathrm{e}^{-vs}}\,\mathrm{d}v=\int_s^{\infty}\frac{\mathrm{e}^{-u}}{1-\mathrm{e}^{-u}}\,\mathrm{d}u$$

$$= \int_s^1 \frac{e^{-u}}{1-e^{-u}}\,du + \int_1^\infty \frac{e^{-u}}{1-e^{-u}}\,du \tag{8.22}$$

$$= \int_s^1 \frac{du}{e^u - 1} + c \leqslant \int_s^1 \frac{du}{u} + c = c - \log s$$

for some constant c. Thus we find that

$$g(s) = I(s) + O(\log(s^{-1})) \quad \text{as } s \to 0^+ . \tag{8.23}$$

Combining (8.23) with (8.20) we see that

$$g(s) \sim \pi^2/(6s) \quad \text{as } s \to 0^+ . \tag{8.24}$$

Therefore, choosing $s = \pi/(6n)^{1/2}$, $x = \exp(-s)$ in Lemma 8.1, we obtain a bound of the form

$$p(n) \leqslant \exp((1 + o(1))\pi(2/3)^{1/2}n^{1/2}) \quad \text{as } n \to \infty . \tag{8.25}$$

⊠

Lemma 8.1 yields a lower bound for $n!$ that is only a factor of about $n^{1/2}$ away from optimal. That is common. Usually, when the function $f(z)$ is reasonably smooth, the best bound obtainable from Lemma 8.1 will only be off from the correct value by a polynomial factor of n, and often only by a factor of $n^{1/2}$.

The estimate of Lemma 8.1 can often be improved with some additional knowledge about the f_n. For example, if $f_{n+1} \geqslant f_n$ for all $n \geqslant 0$, then we have

$$x^{-n}f(x) \geqslant f_n + f_{n+1}x + f_{n+2}x^2 + \cdots \geqslant f_n(1-x)^{-1} . \tag{8.26}$$

For $f_n = p(n)$, the partition function, this yields an upper bound for $p(n)$ that is too large by a factor of $n^{1/4}$.

To optimize the bound of Lemma 8.1, one should choose $x \in (0, R)$ carefully. Usually there is a single best choice. In some pathological cases the optimal choice is obtained by letting $x \to 0^+$ or $x \to R^-$. However, usually we have $\lim_{x \to R^-} f(x) = \infty$, and $[z^m]f(z) > 0$ for some m with $0 \leqslant m < n$ as well as for some $m > n$. Under these conditions it is easy to see that

$$\lim_{x \to 0^+} x^{-n}f(x) = \lim_{x \to R^-} x^{-n}f(x) = \infty . \tag{8.27}$$

Thus it does not pay to make x too small or too large. Let us now consider

$$g(x) = \log(x^{-n}f(x)) = \log f(x) - n\log x . \tag{8.28}$$

Then

$$g'(x) = \frac{f'}{f}(x) - \frac{n}{x} , \tag{8.29}$$

and the optimal choice must be at a point where $g'(x) = 0$. For most commonly encountered functions $f(x)$, there exists a constant $x_0 > 0$ such that

$$\left(\frac{f'}{f}\right)'(x) > 0 \tag{8.30}$$

for $x_0 < x < R$, and so $g''(x) > 0$ for all $x \in (0, R)$ if n is large enough. For such n there is then a unique choice of x that minimizes the bound of Lemma 8.4. However, one major advantage of Lemma 8.1 is that its bound holds for all x. To apply this lemma, one can use any x that is convenient to work with. Usually if this choice is not too far from the optimal one, the resulting bound is fairly good.

We have already remarked above that the bound of Lemma 8.1 is usually close to best possible. It is possible to prove general lower bounds that show this for a wide class of functions. The method, originated in Mazo and Odlyzko (1990) and developed in Odlyzko (1992), relies on simple elementary arguments. However, the lower bounds it produces are substantially weaker than the upper bounds of Lemma 8.1. Furthermore, to apply them it is necessary to estimate accurately the minimum of $x^{-n}f(x)$, instead of selecting any convenient values of x. A more general version of the bound below is given in Odlyzko (1992).

Theorem 8.4. *Suppose that* $f(x) = \sum f_n x^n$ *converges for* $|x| < 1$, $f_n \geqslant 0$ *for all* n, $f_{m_0} > 0$ *for some* m_0, *and* $\sum f_n = \infty$. *Then for* $n \geqslant m_0$, *there is a unique* $x_0 = x_0(n) \in (0,1)$ *that minimizes* $x^{-n}f(x)$. *Let* $s_0 = -\log x_0$, *and*

$$A = \frac{\partial^2}{\partial s^2} \log f(\mathrm{e}^{-s})\Big|_{s=s_0} . \tag{8.31}$$

If $A \geqslant 10^6$ *and for all* t *with*

$$s_0 \leqslant t \leqslant s_0 + 20A^{-1/2} \tag{8.32}$$

we have

$$\left| \frac{\partial^3}{\partial s^3} \log f(\mathrm{e}^{-s})\Big|_{s=t} \right| \leqslant 10^{-3}A^{3/2} , \tag{8.33}$$

then

$$\sum_{k=0}^{n} f_k \geqslant x_0^{-n} f(x_0) \exp(-30 s_0 A^{1/2} - 100) . \tag{8.34}$$

As is usual for Tauberian theorems, Theorem 8.4 only provides bounds on the sum of coefficients of $f(z)$. As we mentioned before, this is unavoidable when one relies only on information about the behavior of $f(z)$ for z a positive real number. The conditions that Theorem 8.4 imposes on the derivatives are usually satisfied in combinatorial enumeration applications and are easy to verify.

Example 8.5 (*Lower bound for the partition function*). Let $f(z)$ and $g(s)$ be as in Example 8.3. We showed there that $g(s)$ satisfies (8.24) and similar rough estimates show that $g'(s) \sim -\pi^2/(6s^2)$, $g''(s) \sim \pi^2/(3s^3)$, and $g'''(s) \sim -\pi^2/s^4$ as $s \to 0^+$. Therefore the hypotheses of Theorem 8.4 are satisfied, and we obtain a lower bound for $p(0) + p(1) + \cdots + p(n)$. If we only use the estimate (8.24) for $g(s)$, then we can only conclude that for $x = \mathrm{e}^{-s}$,

$$\log(x^{-n}f(x)) = ns + g(s) \sim ns + \pi^2/(6s) \quad \text{as } s \to 0 , \tag{8.35}$$

and so the minimum value occurs at $s \sim \pi/(6n)^{1/2}$ as $n \to \infty$. This only allo to conclude that for every $\varepsilon > 0$ and n large enough,

$$\log(p(0) + \cdots + p(n)) \geqslant (1 - \varepsilon)\pi(2/3)^{1/2}n^{1/2} . \tag{8.36}$$

However, we can also conclude even without further computations that this lower bound will be within a multiplicative factor of $\exp(cn^{1/4})$ of the best upper bound that can be obtained from Lemma 8.1 for some $c > 0$ (and therefore within a multiplicative factor of $\exp(cn^{1/4})$ of the correct value). In particular, if we use the estimate (8.14) for $g(s)$, we find that for some $c' > 0$,

$$p(0) + \cdots + p(n) \geqslant \exp(\pi(2/3)^{1/2}n^{1/2} - c'n^{1/4}) . \tag{8.37}$$

Since $p(k) \leqslant p(k+1)$, the quantity on the right-hand side of (8.37) is also a lower bound for $p(n)$ if we increase c', since $(n+1)p(n) \geqslant p(0) + \cdots + p(n)$. ⊠

The differencing trick described at the introduction to section 8 could also be used to estimate $p(n)$, since Theorem 8.1 can be applied to the generating function of $p(n+1) - p(n)$. However, since the error term is a multiplicative factor of $\exp(cn^{1/4})$, it is simpler to use the approach above, which bounds $p(n)$ below by $(p(0) + \cdots + p(n))/(n+1)$.

Brigham (1950) has proved a general theorem about asymptotics of partition functions that can be derived from Theorem 8.4. [For other results and references for partition asymptotics, see Andrews (1976), Ayoub (1963), and Fristedt (1993).]

Theorem 8.6. *Suppose that*

$$f(z) = \prod_{k=1}^{\infty}(1 - z^k)^{-b(k)} = \sum_{n=0}^{\infty} a(n)z^n , \tag{8.38}$$

where the $b(k) \in \mathbb{Z}$, $b(k) \geqslant 0$ *for all* k, *and that for some* $C > 0$, $u > 0$, *we have*

$$\sum_{k \leqslant x} b(k) \sim Cx^u(\log x)^v \quad \text{as } x \to \infty . \tag{8.39}$$

Then

$$\log\left(\textstyle\sum_{n\leqslant m} a(n)\right) \sim u^{-1}\{Cu\Gamma(u+2)\zeta(u+1)\}^{1/(u+1)} \cdot (u+1)^{(u-v)/(u+1)} m^{u/(u+1)} (\log m)^{v/(u+1)} \tag{8.40}$$

as $m \to \infty$.

If $b(k) = 1$ for all k, $a(n)$ is p_n, the ordinary partition function. If $b(k) = k$ for all k, $a(n)$ is the number of plane partitions of n. Thus Brigham's theorem covers a wide class of interesting partition functions. The cost of this generality is that we obtain only the asymptotics of the logarithm of the summatory function of the partitions being enumerated. [For better estimates of the number of plane partitions, for example, see Almkvist (1993), Gordon and Houten (1969), and Wright (1931). For ordinary partitions, we have the expansion (1.3).]

Brigham's proof of Theorem 8.6 first shows that

$$f(e^{-w}) \sim Cw^{-u}(-\log w)^{v}\Gamma(u+1)\zeta(u+1) \quad \text{as } w \to 0^{+} \tag{8.41}$$

and then invokes the Hardy–Ramanujan Tauberian theorem (Rademacher 1937). Instead, one can obtain a proof from Theorem 8.4. The advantage of using Theorem 8.4 is that it is much easier to generalize. Hardy and Ramanujan proved their Tauberian theorem only for functions whose growth rates are of the form given by (8.41). Their approach can be extended to other functions, but this is complicated to do. In contrast, Theorem 8.4 is easy to apply. The conditions of Theorem 8.4 on the derivatives are not restrictive. For a function $f(z)$ defined by (8.38) we have $B \to \infty$ if $\sum b(k) = \infty$, and the condition (8.33) can be shown to hold whenever there are constants c_1 and c_2 such that for all $w > 1$, and all sufficiently large m,

$$\sum_{k \leqslant mw} b(k) \leqslant c_1 w^{c_2} \sum_{k \leqslant m} b(k) , \tag{8.42}$$

say. The main difficulty in applying Theorem 8.4 to generalizations of Brigham's theorem is in accurately estimating the minimal value in Lemma 8.1.

There are many other applications of Lemma 8.1 and Theorem 8.4. For example, they can be used to prove the results of Gardy and Solé (1992) on volumes of spheres in the Lee metric.

Lemma 8.1 can be generalized in a straightforward way to multivariate generating functions. If

$$f(x,y) = \sum_{m,n \geqslant 0} a_{m,n} x^m y^n \tag{8.43}$$

and $a_{m,n} \geqslant 0$ for all m and n, then for any $x, y > 0$ for which the sum in (8.43) converges we have

$$a_{m,n} \leqslant x^{-m} y^{-n} f(x,y) . \tag{8.44}$$

Generalizations of the lower bound of Theorem 8.4 to multivariate functions can also be derived, but are again harder than the upper bound (Moews 1995).

8.2. *Tauberian theorems*

The Brigham Tauberian theorem for partitions (Brigham 1950), based on the Hardy–Ramanujan Tauberian theorem (Rademacher 1937), was quoted above in section 8.1. It applies to certain generating functions that have (in notation to be introduced in section 10) a large singularity and gives estimates only for the logarithm of the summatory function of the coefficients. Another theorem that is often more precise, but is again designed to deal with rapidly growing partition functions, is that of Ingham (1941), and will be discussed at the end of this section. Most of the Tauberian theorems in the literature apply to functions with small singularities (i.e., ones that do not grow rapidly as the argument approaches the circle of convergence) and give asymptotic relations for the sum of coefficients.

References for Tauberian theorems are Feller (1968, 1971), Ganelius (1971), Hardy (1949), Ingham (1941), and Postnikov (1980). Their main advantage is generality and ease of use, as is shown by the applications made to 0–1 laws in Compton (1987, 1988, 1989). They can often be applied when the information about generating functions is insufficient to use the methods of sections 11 and 12. This is especially true when the circle inside which the generating function converges is a natural boundary beyond which the function cannot be continued.

One Tauberian theorem that is often used in combinatorial enumeration is that of Hardy, Littlewood, and Karamata. We say a function $L(t)$ varies slowly at infinity if, for every $u > 0$, $L(ut) \sim L(t)$ as $t \to \infty$.

Theorem 8.7. *Suppose that $a_k \geqslant 0$ for all k, and that*

$$f(x) = \sum_{k=0}^{\infty} a_k x^k$$

converges for $0 \leqslant x < r$. If there is a $\rho \geqslant 0$ and a function $L(t)$ that varies slowly at infinity such that

$$f(x) \sim (r - x)^{-\rho} L\left(\frac{1}{r - x}\right) \quad \text{as } x \to r^- , \tag{8.45}$$

then

$$\sum_{k=0}^{n} a_k r^k \sim (n/r)^\rho L(n)/\Gamma(\rho + 1) \quad \text{as } n \to \infty . \tag{8.46}$$

Example 8.8 (*Cycles of permutations,* Bender 1974). If S is a set of positive integers, and f_n the probability that a random permutation on n letters will have all cycle lengths in S (i.e., $f_n = a_n/n!$, where a_n is the number of permutations with cycle length in S), then

$$f(z) = \sum_{n=0}^{\infty} f_n z^n = \prod_{k \in S} \exp(z^k/k) = (1 - z)^{-1} \prod_{k \notin S} \exp(-z^k/k) . \tag{8.47}$$

If $|\mathbb{Z}^+ \setminus S| < \infty$, then the methods of sections 10.2 and 11 apply easily, and one finds that

$$f_n \sim \exp\left(-\sum_{k \notin S} 1/k\right) \quad \text{as } n \to \infty . \tag{8.48}$$

This estimate can also be proved to apply for $|\mathbb{Z}^+ \setminus S| = \infty$, provided $|\{1, \ldots, m\} \setminus S|$ does not grow too rapidly when $m \to \infty$. If $|S| < \infty$ (or when $|\{1, \ldots, m\} \cap S|$ does not grow rapidly), the methods of section 12 apply. When $S = \{1, 2\}$, one obtains, for example, the result of Moser and Wyman (1955) that the number of permutations of order 2 is

$$\sim (n/\mathrm{e})^{n/2} 2^{-1/2} \exp(n^{1/2} - 1/4) \quad \text{as } n \to \infty . \tag{8.49}$$

[For sharper and more general results, see Moser and Wyman (1955) and Wilf (1986).] The methods used in these cases are different from the ones we are considering in this section.

We now consider an intermediate case, with

$$|\{1,\dots,m\}\cap S| \sim \rho m \quad \text{as } m \to \infty\,, \tag{8.50}$$

for some fixed ρ, $0 \leqslant \rho \leqslant 1$. This case can be handled by Tauberian techniques. To apply Theorem 8.7, we need to show that $L(t) = f(1-t^{-1})t^{-\rho}$ varies slowly at infinity. This is equivalent to showing that for any $u \in (0,1)$,

$$f(1-t^{-1}) \sim f(1-t^{-1}u)u^{\rho} \quad \text{as } t \to \infty\,. \tag{8.51}$$

Because of (8.47), it suffices to prove that

$$\sum_{k\in S} k^{-1}\{(1-t^{-1})^k - (1-t^{-1}u)^k\} = \rho \log u + \mathrm{o}(1) \quad \text{as } t \to \infty\,, \tag{8.52}$$

but this is easy to deduce from (8.50) using summation by parts (section 5). Therefore we find from Theorem 8.7 that

$$\sum_{n=0}^{m} f_n \sim f(1-1/n)\Gamma(\rho+1)^{-1} \quad \text{as } n \to \infty\,. \tag{8.53}$$

[For additional results and references on this problem see Pavlov (1992).] ☒

As the above example shows, Tauberian theorems yield estimates under weak assumptions. These theorems do have some disadvantages. Not only do they usually estimate only the summatory function of the coefficients, but they normally give no bounds for the error term. [See Ganelius (1971) for some Tauberian theorems with remainder terms.] Furthermore, they usually apply only to functions with nonnegative coefficients. Sometimes, as in the following theorem of Hardy and Littlewood, one can relax the nonnegativity condition slightly.

Theorem 8.9. *Suppose that $a_k \geqslant -c/k$ for some $c > 0$,*

$$f(z) = \sum_{k=1}^{\infty} a_k x^k\,, \tag{8.54}$$

and that $f(x)$ converges for $0 < x < 1$, and that

$$\lim_{x\to 1-} f(x) = A\,. \tag{8.55}$$

Then

$$\lim_{n\to\infty} \sum_{k=1}^{n} a_k = A\,. \tag{8.56}$$

Some condition such as $a_k \geqslant -c/k$ on the a_k is necessary, or otherwise the theorem would not hold. For example, the function

$$f(x) = \frac{1-x}{1+x} = 1 - 2x + 2x^2 - \cdots \tag{8.57}$$

satisfies (8.55) with $A = 0$, but (8.56) fails.

We next present an example that shows an application of the above results in combination with other asymptotic methods that were presented before.

Example 8.10 (*Permutations with distinct cycle lengths*). The probability that a random permutation on n letters will have cycles of distinct lengths is $[z^n]f(z)$, where

$$f(z) = \prod_{k=1}^{\infty} \left(1 + \frac{z^k}{k}\right) . \tag{8.58}$$

Greene and Knuth (1982) note that this is also the limit as $p \to \infty$ of the probability that a polynomial of degree n factors into irreducible polynomials of distinct degrees modulo a prime p. It is shown in Greene and Knuth (1982) that

$$[z^n]f(z) = \mathrm{e}^{-\gamma}(1 + n^{-1}) + \mathrm{O}(n^{-2}\log n) \quad \text{as } n \to \infty , \tag{8.59}$$

where $\gamma = 0.577\ldots$ is Euler's constant. A simplified version of the argument of Greene and Knuth (1982) will be presented that shows that

$$[z^n]f(z) \sim \mathrm{e}^{-\gamma} \quad \text{as } n \to \infty . \tag{8.60}$$

Methods for obtaining better expansions, even more precise than that of (8.59), are discussed in section 11.2. For related results obtained by probabilistic methods, see Arratia and Tavaré (1992b).

We have, for $|z| < 1$,

$$\begin{aligned} f(z) &= (1+z)\exp\left(\sum_{k=2}^{\infty} \log(1 + z^k/k)\right) \\ &= (1+z)\exp\left(\sum_{k=2}^{\infty} z^k/k + g(z)\right) \\ &= (1+z)(1-z)^{-1}\exp(g(z)) , \end{aligned} \tag{8.61}$$

where

$$g(z) = -z + \sum_{m=2}^{\infty} \frac{(-1)^{m-1}}{m} \sum_{k=2}^{\infty} \frac{z^{mk}}{k^m} . \tag{8.62}$$

Since the coefficients of $g(z)$ are small, the double sum in (8.62) converges for $z = 1$, and we have

$$\begin{aligned} g(1) = \lim_{z \to 1-} g(z) &= -1 + \sum_{k=2}^{\infty} \sum_{m=2}^{\infty} \frac{(-1)^{m-1}}{m} k^{-m} \\ &= -1 + \sum_{k=2}^{\infty} \{\log(1 + k^{-1}) - k^{-1}\} \\ &= -\log 2 + \lim_{n \to \infty} (\log(n+1) - H_n) = -\log 2 - \gamma \ , \end{aligned} \tag{8.63}$$

where $H_n = 1 + 1/2 + 1/3 + \cdots + 1/n$ is the nth *harmonic number*. Therefore, by (8.61), we find from Theorem 8.9 that if $f_n = [z^n]f(z)$, then

$$f_0 + f_1 + \cdots + f_n \sim n\,\mathrm{e}^{-\gamma} \quad \text{as } n \to \infty \ . \tag{8.64}$$

To obtain asymptotics of f_n, we note that if $h_n = [z^n]\exp(g(z))$, then by (8.61),

$$f_n = 2h_0 + 2h_1 + \cdots + 2h_{n-1} + h_n \ . \tag{8.65}$$

We next obtain an upper bound for $|h_n|$. There are several ways to proceed. The method used below gives the best possible result $|h_n| = \mathrm{O}(n^{-2})$.

Since $g(z)$ has the power-series expansion (8.62), and $h_n = [z^n]\exp(g(z))$, comparison of terms in the full expansion of $\exp(g(z))$ and $\exp(v(z))$ shows that $|h_n| \leqslant [z^n]\exp(v(z))$, where $v(z)$ is any power series such that $|[z^n]g(z)| \leqslant [z^n]v(z)$. For $n \geqslant 2$,

$$[z^n]g(z) = \sum_{\substack{m|n \\ m \geqslant 2 \\ m < n}} \frac{(-1)^{m-1}}{m} \left(\frac{m}{n}\right)^m \ . \tag{8.66}$$

The term $(m/n)^m$ is monotone decreasing for $1 \leqslant m \leqslant n/\mathrm{e}$, since its derivative with respect to m is $\leqslant 0$ in that range. Therefore

$$|[z^n]g(z)| \leqslant \frac{1}{2}\left(\frac{2}{n}\right)^2 + \sum_{3 \leqslant m \leqslant n/3} \frac{1}{m}\left(\frac{3}{n}\right)^3 + \frac{2}{n} 2^{-n/2} \leqslant 10 n^{-2} \ , \tag{8.67}$$

say. Hence we can take

$$v(z) = 10 \sum_{n=1}^{\infty} n^{-2} z^n \ , \tag{8.68}$$

and then we need to estimate

$$w_n = [z^n]\exp(v(z)) \ . \tag{8.69}$$

We let $w(z) = \exp(v(z))$, and note that

$$w'(z) = v'(z) w(z) \ ; \tag{8.70}$$

so for $n \geqslant 1$,

$$nw_n = 10\sum_{k=0}^{n-1} w_k(n-k)^{-1} . \tag{8.71}$$

Further, since $v(1) < \infty$, and $w_n \geqslant 0$ for all n, we have $w_n \leqslant A = w(1) = \exp(v(1))$ for all n. Let $B = 10^6 A$ and note that $w_n \leqslant Bn^{-2}$ for $1 \leqslant n \leqslant 10^3$. Suppose now that $w_m \leqslant Bm^{-2}$ for $1 \leqslant m < n$ for some $n \geqslant 10^3$. We will prove that $w_n \leqslant Bn^{-2}$, and then by induction this inequality will hold for all $n \geqslant 1$. We apply eq. (8.70). For $0 \leqslant k \leqslant 100$, we use $w_k \leqslant A$, $(n-k)^{-1} \leqslant 2n^{-1}$. For $100 < k \leqslant n/2$,

$$w_k(n-k)^{-1} \leqslant Bk^{-2}(n-k)^{-1} \leqslant 2Bk^{-2}n^{-1} , \tag{8.72}$$

and so

$$\sum_{100 \leqslant k \leqslant n/2} w_k(n-k)^{-1} \leqslant B(40n)^{-1} . \tag{8.73}$$

Finally,

$$\sum_{n/2 < k \leqslant n-1} w_k(n-k)^{-1} \leqslant 4Bn^{-2} \sum_{n/2 < k \leqslant n-1} (n-k)^{-1} \leqslant 4Bn^{-2}H_n . \tag{8.74}$$

Therefore, by (8.71),

$$nw_n \leqslant 2000An^{-1} + B(4n)^{-1} + 4BH_n n^{-2} \leqslant Bn^{-1} , \tag{8.75}$$

which completes the induction step and proves that $w_n \leqslant Bn^{-2}$ for all $n \geqslant 1$. ☒

There are Tauberian theorems that apply to generating functions with rapidly growing coefficients but are more precise than Brigham's theorem or the estimates obtainable with the methods of section 8.1. One of the most useful is Ingham's Tauberian theorem for partitions (Ingham 1941). The following result is a corollary of the more general Theorem 2 of Ingham (1941).

Theorem 8.11. *Let* $1 \leqslant u_1 < u_2 < \ldots$ *be positive integers such that*

$$|\{u_j\colon u_j \leqslant x\}| = Bx^{\beta} + R(x) , \tag{8.76}$$

where $B > 0$, $\beta > 0$, *and*

$$\int_1^y x^{-1}R(x)\,\mathrm{d}x = b\log y + c + \mathrm{o}(1) \quad \textit{as } y \to \infty . \tag{8.77}$$

Let

$$a(z) = \sum_{n=1}^{\infty} a_n z^n = \prod_{j=1}^{\infty} (1 - z^{u_j})^{-1} , \tag{8.78}$$

$$a^*(z) = \sum_{n=1}^{\infty} a_n^* z^n = \prod_{j=1}^{\infty} (1 + z^{u_j}) . \tag{8.79}$$

Then, as $m \to \infty$,

$$\sum_{n=1}^{m} a_n \sim (2\pi)^{-1/2}(1-\alpha)^{1/2} e^c V^{-\alpha(b+1/2)} m^{(b+1/2)(1-\alpha)-1/2} \exp(\alpha^{-1}(Vm)^{\alpha}) \,, \tag{8.80}$$

$$\sum_{n=1}^{m} a_n^* \sim (2\pi)^{-1/2}(1-\alpha)^{1/2} 2^b (V^* m)^{-\alpha/2} \exp(\alpha^{-1}(V^* m)^{\alpha}) \,, \tag{8.81}$$

where

$$\alpha = \beta(\beta+1)^{-1}, \quad V = \{B\beta\Gamma(\beta+1)\zeta(\beta+1)\}^{1/\beta} \,, \quad V^* = (1-2^{-\beta})^{1/\beta} V \,. \tag{8.82}$$

If $u_1 = 1$, *then as* $n \to \infty$

$$a_n \sim (2\pi)^{-1/2}(1-\alpha)^{1/2} e^c V^{-\alpha(b-1/2)} n^{(b-1/2)(1-\alpha)-1/2} \exp(\alpha^{-1}(Vn)^{\alpha}) \,, \tag{8.83}$$

and if $1, 2, 4, 8, \ldots$ *all belong to* $\{u_j\}$, *then*

$$a_n^* \sim (2\pi)^{-1/2}(1-\alpha)^{1/2} 2^b (V^*)^{\alpha/2} n^{\alpha/2-1} \exp(\alpha^{-1}(V^* n)^{\alpha}) \,. \tag{8.84}$$

Theorem 8.11 provides more precise information than Brigham's Theorem 8.6, but under more restrictive conditions. It is derived from Ingham's main result, Theorem 1 of Ingham (1941), which can be applied to wider classes of functions. However, that theorem cannot be used to derive Theorem 8.6. The disadvantage of Ingham's main theorem is that it requires knowledge of the behavior of the generating function in the complex plane, not just on the real axis. On the other hand, the region where this behavior has to be known is much smaller than it is for the analytic methods that give more accurate answers, and which are presented in sections 10–12. Only behavior of the generating functions $\Pi(1 - z^{\lambda_j})^{-1}$ or $\Pi(1 + z^{\lambda_j})$ in an angle $|\mathrm{Arg}|(1-z)| \leqslant \pi/2 - \delta$ for some $\delta > 0$ needs to be controlled.

Ingham's paper (1941) contains an extended discussion of the relations between different Tauberian theorems and of the necessity for various conditions.

9. Recurrences

This section presents some basic methods for handling recurrences. The title is slightly misleading, since almost all of this chapter is devoted to methods that are useful in this area. Almost all asymptotic estimation problems concern quantities that are defined through implicit or explicit recurrences. Furthermore, the most common and most effective method of solving recurrences is often to determine their generating functions and then apply the methods presented in the other sections. However, there are many recurrences, and those discussed in sections 9.4 and 9.5 require special methods that do not fit into other sections. These methods deserve to be included, so it seems preferable to explain them after treating some of the more common types of recurrences, even though those could have been covered elsewhere in this chapter.

Since generating functions are the most powerful tool for handling combinatorial recurrences, all the books listed in section 18 that help in dealing with combinatorial identities and generating functions are also useful in handling recurrences. Methods for recurrences that are not amenable to generating function methods are presented in Graham et al. (1989) and Greene and Knuth (1982). Lueker (1980) is an introductory survey to some recurrence methods.

Wimp's (1984) book is concerned primarily with numerical stability problems in computing with recurrences. Such problems are important in computing values of orthogonal polynomials, for example, but seldom arise in combinatorial enumeration. However, there are sections of Wimp (1984) that are relevant to our topic, for example to the discussion of differential equations in section 9.2.

9.1. Linear recurrences with constant coefficients

The most famous sequence that satisfies a linear recurrence with constant coefficients is that of the Fibonacci numbers, defined by $F_0 = F_1 = 1$, $F_n = F_{n-1} + F_{n-2}$ for $n \geqslant 2$. There are many others that are only slightly less well known. Fortunately, the theory of such sequences is well developed, and from the standpoint of asymptotic enumeration their behavior is well understood. [For a survey of number-theoretic results, together with a list of many unsolved problems about such sequences that arise in that area, see Cerlienco et al. (1987).] There are even several different approaches to solving linear recurrences with constant coefficients. The one we emphasize here is that of generating functions, since it fits in best with the rest of this chapter. For other approaches, see Milne-Thomson (1933) or Nörlund (1924), for example.

Suppose that we have a linear recurrence or a system of recurrences and have found that the generating function $f(z)$ we are interested in has the form

$$f(z) = \frac{G(z)}{h(z)}, \tag{9.1}$$

where $G(z)$ and $h(z)$ are polynomials. The basic tool for obtaining asymptotic information about $[z^n]f(z)$ is the partial fraction expansion of a rational function (Henrici 1974–86). Dividing $G(z)$ by $h(z)$ we obtain

$$f(z) = p(z) + \frac{g(z)}{h(z)}, \tag{9.2}$$

where $p(z)$, $g(z)$, and $h(z)$ are all polynomials in z and $\deg g(z) < \deg h(z)$. We can assume that $h(0) \neq 0$, since if that were not the case, we would have $g(0) = 0$ (as in the opposite case $f(z)$ would not be a power series in z, but would have terms such as z^{-1} or z^{-2}) and we could cancel a common factor of z from $g(z)$ and $h(z)$. Therefore, if $d = \deg h(z)$, we can write

$$h(z) = h(0) \prod_{j=1}^{d'} \left(1 - \frac{z}{z_j}\right)^{m_j} , \tag{9.3}$$

where z_j, $1 \leqslant j \leqslant d'$ are the distinct roots of $h(z) = 0$, z_j has multiplicity $m_j \geqslant 1$, and $\sum m_j = d$. Hence we find (Graham et al. 1989, Henrici 1974–86) that for certain constants $c_{j,k}$,

$$\begin{aligned} f(z) &= p(z) + \sum_{j=1}^{d'} \sum_{k=1}^{m_j} \frac{c_{j,k}}{(1 - z/z_j)^k} \\ &= p(z) + \sum_{j=1}^{d'} \sum_{k=1}^{m_j} c_{j,k} \sum_{h=0}^{\infty} \binom{h+k-1}{k-1} z^h z_j^{-h} . \end{aligned} \tag{9.4}$$

Thus

$$a_n = [z^n]p(z) + \sum_{j=1}^{d'} \sum_{k=1}^{m_j} c_{j,k} \binom{h+k-1}{k-1} z_j^{-n} . \tag{9.5}$$

When $m_j = 1$,

$$c_{j,1} = \frac{-g(z_j)}{z_j h'(z_j)} , \tag{9.6}$$

and explicit formulas for the $c_{j,k}$ when $m_j > 1$ can also be derived (Graham et al. 1989), but they are unwieldy and seldom used.

Example 9.1 (*Fibonacci numbers*). As was noted in Example 6.3,

$$F(z) = \sum_{n=0}^{\infty} F_n z^n = \frac{z}{1 - z - z^2} .$$

Now

$$h(z) = 1 - z - z^2 = (1 + \phi^{-1} z)(1 - \phi z) , \tag{9.7}$$

where $\phi = (1 + 5^{1/2})/2$ is the golden ratio. Therefore

$$F(z) = \frac{1}{\sqrt{5}} \left(\frac{1}{1 - \phi z} - \frac{1}{1 + \phi^{-1} z} \right) \tag{9.8}$$

and for $n \geqslant 0$,

$$F_n = [z^n]F(z) = 5^{-1/2}(\phi^n - (-\phi)^{-n}) . \tag{9.9}$$

⊠

The partial fraction expansion (9.4) shows that the first-order asymptotics of sequence a_n satisfying a linear recurrence of the form (6.30) are determined by the smallest zeros of the characteristic polynomial $h(z)$. The full asymptotic expansion is given by (9.5), and involves all the zeros. In practice, using (9.5) presents some difficulties, in that multiplicities of zeros are not always easy to determine, and the coefficients $c_{j,k}$ are often even harder to deal with. Eventually, for large n, their influence becomes negligible, but when uniform estimates are required they present a problem. In such cases the following theorem is often useful.

Theorem 9.2. *Suppose that $f(z) = g(z)/h(z)$, where $g(z)$ and $h(z)$ are polynomials, $h(0) \neq 0$, $\deg g(z) < \deg h(z)$, and that the only zeros of $h(z)$ in $|z| < R$ are $\rho_1, \ldots, \rho_k$, each of multiplicity* 1. *Suppose further that*

$$\max_{|z|=R} |f(z)| \leqslant W\ , \tag{9.10}$$

and that $R - |\rho_j| \geqslant \delta$ for some $\delta > 0$ and $1 \leqslant j \leqslant k$. Then

$$\left| [z^n]f(z) + \sum_{j=1}^{k} \frac{g(\rho_j)}{h'(\rho_j)} \rho_j^{-n-1} \right| \leqslant WR^{-n} + \delta^{-1}R^{-n} \sum_{j=1}^{k} |g(\rho_j)/h'(\rho_j)|\ . \tag{9.11}$$

Theorem 9.2 is derived using methods of complex variables, and a proof is sketched in section 10. That section also discusses how to locate all the zeros $\rho_1, \ldots, \rho_k$ of a polynomial $h(z)$ in a disk $|z| < R$. In general, the zero location problem is not a serious one in enumeration problems. Usually there is a single positive real zero that is closer to the origin than any other, it can be located accurately by simple methods, and R is chosen so that $|z| < R$ encloses only that zero.

Example 9.3 (*Sequences with forbidden subblocks*). We continue with the problem presented in Examples 6.4 and 6.8. Both $F_A(z)$ and $G_A(z)$ have as denominators

$$h(z) = z^k + (1 - 2z)C_A(z)\ , \tag{9.12}$$

which is a polynomial of degree exactly k. Later, in Example 10.11, we will show that for $k \geqslant 9$, $h(z)$ has exactly one zero ρ in $|z| \leqslant 0.6$, and that for $|z| = 0.55$, $|h(z)| \geqslant 1/100$. Furthermore, by Example 6.7, $\rho \to 1/2$ as $k \to \infty$. On $|z| = 0.55$,

$$|F_A(z)| \leqslant 100 \cdot (0.55)^k\ . \tag{9.13}$$

Theorem 9.2 then shows, for example, that for $n > k \geqslant k_0$,

$$\begin{aligned} \left| [z^n]F_A(z) + \frac{C_A(\rho)\rho^{-n-1}}{h'(\rho)} \right| &\leqslant 100(0.55)^{k\ \ n} + 40(0.55)^{\ n}|h'(\rho)|^{\ 1} \\ &\leqslant 50(0.55)^{-n}\ , \end{aligned} \tag{9.14}$$

since by Example 6.7, as $k \to \infty$,

$$h'(\rho) = k\rho^{k-1} - 2C_A(\rho) + (1-2\rho)C_A'(\rho) \sim -2C_A(\rho) \sim -\rho^{-1} . \tag{9.15}$$

The estimate (9.14), when combined with the expansions of Example 6.7, gives accurate approximations for p_n, the probability that A does not appear as a block among the first n coin tosses. We have

$$\begin{aligned} p_n &= 2^{-n}[z^n]F_z(z) \\ &= -2^{-n}C_A(\rho)\rho^{-n-1}(h'(\rho))^{-1} + O(\exp(-0.09n)) . \end{aligned} \tag{9.16}$$

We now estimate $h'(\rho)$ as before, in (9.15), but more carefully, putting in the approximation for ρ from Example 6.7. We find that

$$h'(\rho) = -\rho^{-1} + O(k2^{-k}) , \tag{9.17}$$

and

$$\rho^{-n} = 2^n \exp(-n(2^kC_A(1/2))^{-1} + O(nk2^{-2k})) . \tag{9.18}$$

Therefore

$$p_n = \exp(-n(2^kC_A(1/2))^{-1} + O(nk2^{-2k})) + O(\exp(-n/12)) . \tag{9.19}$$

This shows that p_n has a sharp transition. It is close to 1 for $n = o(2^k)$, and then, as n increases through 2^k, drops rapidly to 0. (The behavior on the two sides of 2^k is not symmetric, as the drop towards 0 beyond 2^k is much faster than the increases towards 1 on the other side.) For further results and applications of such estimates, see Guibas and Odlyzko (1978, 1980). Estimates such as (9.19) yield results sharper than those of Example 6.8. They also prove (see Example 14.1) that the expected lengths of the longest run of 0's in a random sequence of length n is $\log_2 n + u(\log_2 n) + o(1)$ as $n \to \infty$, where $u(x)$ is a continuous function that is not constant and satisfies $u(x+1) = u(x)$. [See also the discussion of carry propagation in Knuth (1981).] For other methods and results in this area, see Arratia et al. (1990b). ⊠

Inhomogeneous recurrences with constant coefficients, say,

$$a_n = \sum_{i=1}^{d} c_i a_{n-i} + b_n, \quad n \geqslant d , \tag{9.20}$$

are not covered by the techniques discussed above. One can still use the basic-generating function approach to derive the ordinary generating function of a_n, but this time it is in terms of the ordinary generating function of b_n. If b_n does not grow too rapidly, the "subtraction of singularities" method of section 10.2 can be used to derive the asymptotics of a_n in a form similar to that given by (9.26).

9.2. Linear recurrences with varying coefficients

Linear recurrences with constant coefficients have a nice and complete theory. That is no longer the case when one allows coefficients that vary with the index. This is not a fault of mathematicians in not working hard enough to derive elegant results, but reflects the much more complicated behavior that can occur. The simplest case is when the recurrence has a finite number of terms, and the coefficients are polynomials in n.

Example 9.4 (*Two-sided generalized Fibonacci sequences*). Let t_n be the number of integer sequences $(b_j, \ldots, b_2, b_1, 1, 1, a_1, a_2, \ldots, a_k)$ with $j+k+2=n$ in which each b_i is the sum of one or more contiguous terms immediately to its right, and each a_i is likewise the sum of one or more contiguous terms immediately to its left. It was shown in Fishburn et al. (1989) that $t_1 = t_2 = 1$ and that

$$t_{n+1} = 2nt_n - (n-1)^2 t_{n-1} \quad \text{for } n \geqslant 2\,. \tag{9.21}$$

If we let

$$t(z) = \sum_{n=1}^{\infty} \frac{t_n z^{n-1}}{(n-1)!} \tag{9.22}$$

be a modified exponential generating function, then the recurrence (9.21) shows that

$$t'(z)(1-z)^2 - t(z)(2-z) = 1\,. \tag{9.23}$$

Standard methods for solving ordinary differential equations, together with the initial conditions $t_1 = t_2 = 1$, then yield the explicit solution

$$t(z) = (1-z)^{-1} \exp((1-z)^{-1}) \left[C + \int_z^1 (1-w)^{-1} \exp(-(1-w)^{-1})\, \mathrm{d}w \right], \tag{9.24}$$

where

$$C = \mathrm{e}^{-1} - \int_0^1 (1-w)^{-1} \exp(-(1-w)^{-1})\, \mathrm{d}w = 0.148495\ldots\,. \tag{9.25}$$

Once the explicit formula (9.24) for $t(z)$ is obtained, the methods of section 12 give the estimate

$$t_n \sim C(n-1)!(\mathrm{e}/\pi)^{1/2} \exp(2n^{1/2})(2n^{1/4})^{-1} \quad \text{as } n \to \infty\,. \tag{9.26}$$

It is easy to show that the absolute value of

$$(1-z)^{-1} \exp((1-z)^{-1}) \int_z^1 (1-w)^{-1} \exp(-(1-w)^{-1})\, \mathrm{d}w \tag{9.27}$$

is small for $|z| < 1$. Therefore the asymptotics of the t_n are determined by the behavior of coefficients of

$$C(1-z)^{-1}\exp((1-z)^{-1}) , \tag{9.28}$$

and that can be obtained easily. The estimate (9.26) then follows. ☒

To see just how different the behavior of linear recurrences with polynomial coefficients can be from those with constant coefficients, compare the behavior of the sequences in Example 9.4 above and Example 9.5 (given below). The existence of such differences should not be too surprising, since after all even the first-order recurrence $a_n = na_{n-1}$ for $n \geqslant 2$, $a_1 = 1$, has the obvious solution $a_n = n!$, which is not at all like the solutions to constant coefficient-recurrences. However, when $a_n = na_{n-1}$, a simple change of variables, namely $a_n = b_n n!$, transforms this recurrence into the trivial one of $b_n = b_{n-1} = \cdots = b_1 = 1$ for all n. Such rescaling is among the most fruitful methods for dealing with nonlinear recurrences, even though it is seldom as simple as for $a_n = n!$.

Example 9.4 is typical in that a sequence satisfying a linear recurrence of the form

$$a_n = \sum_{j=1}^{r} c_j(n) a_{n-j} , \quad n \geqslant r , \tag{9.29}$$

where r is fixed and the $c_j(n)$ are rational functions (a P-recursive sequence in the notation of section 6.3) can always be transformed into a differential equation for a generating function. Whether anything can be done with that generating function depends strongly on the recurrence and the form of the generating function. Example 9.4 is atypical in that there is an explicit solution to the differential equation. Further, this explicit solution is a nice analytic function. This is due to the special choice of the form of the generating function. An exponential generating function seems natural to use in that example, since the recurrence (9.21) shows immediately that $t_n \leqslant (2n-2)(2n-4)\cdots 2 = 2^{n-1}(n-1)!$, and a slightly more involved induction proves that t_n grows at least as fast as a factorial. If we tried to use an ordinary generating function

$$u(z) = \sum_{n=1}^{\infty} t_n z^n , \tag{9.30}$$

then the recurrence (9.21) would yield the differential equation

$$z^4 u''(z) + z^3 u'(z) + (1-2z^2)u(z) = z - z^2 , \tag{9.31}$$

which is not as tractable. (This was to be expected, since $u(z)$ is only a formal power series.) Even when a good choice of generating function does yield an analytic function, the differential equation that results may be hard to solve. (One can always find a generating function that is analytic, but the structure of the problem may not be reflected in the resulting differential equation, and there may not be anything nice about it.)

There is an extensive literature on analytic solutions of differential equations (cf. Henrici 1974–86, Hille 1969, 1976, Malgrange 1974, Varadarajan 1991, Wasow 1965), but it is not easy to apply in general. Singularities of analytic functions that satisfy linear differential equations with analytic coefficients are usually of only a few basic forms, and so the methods of sections 11 and 12 suffice to determine the asymptotic behavior of the coefficients. The difficulty is in locating the singularities and determining their nature. We refer to Hille (1969, 1976), Malgrange (1974), Varadarajan (1991), and Wasow (1965) for methods for dealing with this difficulty, since they are involved and so far have been seldom used in combinatorial enumeration. There will be some further discussion of differential equations in section 15.3.

Some aspects of the theory of linear recurrences with constant coefficients do carry over to the case of varying coefficients, even when the coefficients are not rational functions. For example, there will in general be r linearly independent solutions to the recurrence (9.29) (corresponding to the different starting conditions). Also, if a solution a_n has the property that a_{n+1}/a_n tends to a limit α as $n \to \infty$, then $1/\alpha$ is a limit of zeros of

$$1 - \sum_{j=1}^{r} c_j(n) z^j , \tag{9.32}$$

and therefore is often a root of

$$1 - \sum_{j=1}^{r} \left(\lim_{n\to\infty} c_j(n) \right) z^j . \tag{9.33}$$

Whether there are exactly r linearly independent solutions is a difficult problem. Extensive research was done on this topic in the early 20th century (Adams 1928, Batchelder 1927), culminating in the work of Birkhoff and Trjitzinsky (Birkhoff 1911, 1930, Birkhoff and Trjitzinsky 1932, Trjitzinsky 1933a,b). This work applies to recurrences of the form (9.29) where the $c_j(n)$ have Poincaré asymptotic expansions

$$c_j(n) \sim n^{k_j/k} \{ c_{j,0} + c_{j,1} n^{-1/k} + c_{j,2} n^{-2/k} + \cdots \} \quad \text{as } n \to \infty , \tag{9.34}$$

where the k_j and k are integers and $c_{j,0} \neq 0$ if $c_j(n)$ is not identically 0 for all n. It follows from this work that solutions to the recurrence are expressible as linear combinations of elements of the form

$$(n!)^{p/q} \exp(P(n^{1/m})) n^{\alpha} (\log n)^{h} , \tag{9.35}$$

where h, m, p, and q are integers, $P(z)$ a polynomial, and α a complex number. An exposition of this theory and how it applies to enumeration has been given by Wimp and Zeilberger (1985). (There is a slight complication in that most of the literature cited above is concerned with recurrences for functions of a real argument, not sequences, but this is not a major difficulty.) There is still a problem in identifying which linear combination provides the derived solution. Wimp and Zeilberger point out that it is usually easy to show that the largest of the terms of the form (9.35) does show up with a nonzero coefficient, and so determines the asymptotics of a_n up to a multiplicative constant. However, the Birkhoff–Trjitzinsky method does not in general provide any techniques for determining that constant.

The major objection to the use of the Birkhoff–Trjitzinsky results is that they may not be rigorous, since gaps are alleged to exist in the complicated proofs (Immink 1984, Wimp 1991). Furthermore, in almost all combinatorial enumeration applications the coefficients are rational, and so one can use the theory of analytic differential equations.

When there is no way to avoid linear recurrences with coefficients that vary but are not rational, one can sometimes use the work of Kooman (1989, Kooman and Tijdeman 1990), which develops the theory of second-order linear recurrences with almost-constant coefficients.

Example 9.5 (*An oscillating sequence*). Let

$$a_n = \sum_{k=0}^{n} \binom{n}{k} \frac{(-1)^k}{k!}, \quad n = 0, 1, \ldots . \tag{9.36}$$

Then a_n satisfies the linear recurrence

$$a_{n+2} - \left(2 - \frac{2}{n}\right) a_{n+1} + \left(1 - \frac{1}{n}\right) a_n = 0, \quad n \geqslant 0 . \tag{9.37}$$

The methods of Kooman and Tijdeman (1990) can be used to show that for some constants c and ϕ

$$a_n = c n^{-1/4} \sin(2n^{1/2} + \phi) + o(n^{-1/4}) \quad \text{as } n \to \infty , \tag{9.38}$$

which is a much more precise estimate than the crude one mentioned in Example 10.3.

Another, in some ways preferable, method for obtaining asymptotic expansions for a_n is mentioned in Example 12.15. It is based on an explicit form for the generating function of a_n, $f(z) = \sum a_n z^n$. An interchange of orders of summation (easily justified for $|z|$ small, say $|z| < 1/2$) shows that

$$\begin{aligned} f(z) &= \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \sum_{n=k}^{\infty} \binom{n}{k} z^n \\ &= \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \frac{z^k}{(1-z)^{k+1}} = \frac{1}{1-z} \exp\left(-\frac{z}{1-z}\right) . \end{aligned} \tag{9.39}$$

The saddle point method can then be applied to obtain asymptotic expansions for a_n. ⊠

9.3. Linear recurrences in several variables

Linear recurrences in several variables that have constant coefficients can be attacked by methods similar to those used for a single variable. If we have

$$a_{m,n} = \sum_{\substack{i=0 \\ i+j>0}}^{d} \sum_{i=0}^{d} c_{i,j} a_{m-i,n-j} \tag{9.40}$$

for $m, n \geqslant d$, say, then the generating function

$$f(x,y) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} a_{m,n} x^m y^n \tag{9.41}$$

satisfies the relation

$$f(x,y)\left(1 - \sum_{\substack{i=0 \\ i+j>0}}^{d} \sum_{i=0}^{d} c_{i,j} x^i y^j\right) = \sum_{\substack{m=0 \\ m>d \text{ or } n>d}}^{\infty} \sum_{n=0}^{\infty} a_{m,n} x^m y^n$$
$$- \sum_{\substack{i=0 \\ i+j>0}}^{d} \sum_{i=0}^{d} c_{ij} x^i y^j \sum_{\substack{m,n \\ m \leqslant d-i \\ \text{or } n \leqslant d-i}} a_{m,n} x^m y^n \,. \tag{9.42}$$

If $a_{m,n} = 0$ for $0 \leqslant m < d$ and $n \geqslant d$ as well as for $0 \leqslant n < d$ and $m \geqslant d$ (so that all the $a_{m,n}$ are fully determined by $a_{m,n}$ for $0 \leqslant m < d$, $0 \leqslant n < d$), then $f(x,y)$ is a rational function. If this condition does not hold, $f(x,y)$ can be complicated.

The paragraph above shows that under common conditions, constant-coefficient recurrences lead to generating functions that are rational even in several variables. However, even when the rational function is determined, there is no equivalent of partial fraction decomposition to yield elegant asymptotics of the coefficients. Coefficients of multivariate generating functions are much harder to handle than those of univariate functions. There are tools (discussed in section 13), that are usually adequate to handle rational generating functions, but they are not simple.

When the coefficients of the multivariate recurrences vary, available knowledge is extremely limited. Even if the coefficients are polynomials, we obtain a partial differential equation for the generating function. Sometimes there are tricks that lead to a simple solution (cf. Example 15.6), but this is not common.

9.4. Nonlinear recurrences

Nonlinear recurrences come in a great variety of shapes, and the methods that are used to solve them are diverse, depending on the nature of the problem. This section presents a sample of the most useful techniques that have been developed.

Sometimes a nonlinear recurrence has a simple solution because of a nice algebraic factorization. For example, suppose that z_0 is any given complex number, and

$$z_{n+1} = z_n^2 - 2 \quad \text{for } n \geqslant 0 . \tag{9.43}$$

If we set

$$w = (z_0 + (z_0^2 - 4)^{1/2})/2 , \tag{9.44}$$

we have $z_0 = w + w^{-1}$, and more generally

$$z_n = w^{2^n} + w^{-2^n} \quad \text{for } n \geqslant 0 . \tag{9.45}$$

Equation (9.45) is easily established through induction. However, this is an exceptional instance, and already recurrences of the type $z_{n+1} = z_n^2 + c$ for c a complex constant lead to deep questions about the Mandelbrot set and chaotic behavior (Devaney 1989).

Since linear recurrences are well understood, the best that one can hope for when confronted with a nonlinear recurrence is that it might be reducible to a linear one. This works in many situations.

Example 9.6 (*Planted plane trees*). Let $a_{n,h}$ be the number of planted plane trees with n nodes and height $\leqslant h$ (de Bruijn et al. 1972, Greene and Knuth 1982), and let

$$A_h(z) = \sum_{n=0}^{\infty} a_{n,h} z^n . \tag{9.46}$$

Since a tree of height $\leqslant h + 1$ has a root and any number of subtrees, each of height $\leqslant h$,

$$\begin{aligned} A_{h+1}(z) &= z(1 + A_h(z) + A_h(z)^2 + \cdots) \\ &= z(1 - A_h(z))^{-1} . \end{aligned} \tag{9.47}$$

Iterating this recurrence, we obtain a finite continued fraction that looks like

$$A_{h+1}(z) = \frac{z}{1 - \frac{z}{1 - \frac{z}{1 \dots}}} . \tag{9.48}$$

The general theory of continued functions represents a convergent as a quotient of two sequences satisfying recurrences involving the partial quotients. (For references, see Jones and Thron 1980, Perron 1957.) After playing with this idea, one finds that the substitution

$$A_h(z) = \frac{zP_h(z)}{P_{h+1}(z)} \tag{9.49}$$

gives

$$P_{h+1}(z) = P_h(z) - zP_{h-1}(z) , \qquad h \geqslant 2 ,$$

where $P_0(z) = 0$, $P_1(z) = 1$. This is a linear recurrence when we regard z as fixed, and so the theory presented before leads to the explicit representation

$$P_h(z) = (1-4z)^{-1/2}\left\{\left(\frac{1+(1-4z)^{1/2}}{2}\right)^h - \left(\frac{1-(1-4z)^{1/2}}{2}\right)^h\right\} . \tag{9.50}$$

De Bruijn et al. (1972) use this representation to determine the average height of plane trees. ☒

Greene and Knuth (1982, p. 30) note that the continued fraction method of replacing a convergent by a quotient of elements of two sequences in general leads not to a single sequence of polynomials like the $P_h(z)$ of Example 9.6, but to two sequences. This is only slightly harder to handle, and allows one to linearize more complicated recurrences.

There are many additional ways to linearize a recurrence. [A small list is given on p. 31 of Greene and Knuth (1982).] For example, a purely multiplicative relation $a_n = a_{n-1}^2/a_{n-2}$ is transformed into the linear $\log a_n = 2\log a_{n-1} - \log a_{n-2}$ by taking logarithms. One of the most fruitful tricks of this type is taking inverses. Thus $a_n = a_{n-1}/(1+a_{n-1})$ is equivalent to

$$\frac{1}{a_n} = \frac{1}{a_{n-1}} + 1 , \tag{9.51}$$

which has the obvious solution $a_n^{-1} = a_0^{-1} + n$. (This assumes $a_0 \neq -1/k$ for any $k \in \mathbb{Z}^+$.)

Linearization works well, but is limited in applicability. More widely applicable, but producing answers that are not as clear, is approximate linearization, where a given nonlinear recurrence is close to a linear one. The following example combines approximate linearization with bootstrapping.

Example 9.7 (*A quadratic recurrence*). The study of the average height of binary trees in Flajolet and Odlyzko (1982) involves the recurrence

$$a_n = a_{n-1}(1-a_{n-1}) \quad \text{for } n \geqslant 1 , \tag{9.52}$$

with $a_0 = 1/2$. The a_n are monotone decreasing, so we try the inverse trick. We find

$$\frac{1}{a_n} = \frac{1}{a_{n-1}(1-a_{n-1})} = \frac{1}{a_{n-1}} + 1 + \frac{a_{n-1}}{1-a_{n-1}} . \tag{9.53}$$

Iterating this recurrence [but applying it only to the first term on the right-hand side of eq. (9.53)] we obtain

$$\begin{aligned}\frac{1}{a_n} &= \frac{1}{a_{n-2}} + 2 + \frac{a_{n-2}}{1-a_{n-2}} + \frac{a_{n-1}}{1-a_{n-1}} \\ &= \cdots \\ &= \frac{1}{a_0} + n + \sum_{j=0}^{n-1} \frac{a_j}{1-a_j} \\ &= n + 2 + \sum_{j=0}^{n-1} \frac{a_j}{1-a_j} .\end{aligned} \tag{9.54}$$

Equation (9.54) shows that $a_n^{-1} > n$, so $a_n < 1/n$. Applying this bound to a_j for $2 \leqslant j \leqslant n-1$ in the sum on the right-hand side of eq. (9.54), we find that

$$n \leqslant a_n^{-1} \leqslant n + \mathrm{O}(\log n) . \tag{9.55}$$

When we substitute this into (9.54), we find that $a_n^{-1} = n + \log n + \mathrm{o}(\log n)$, and further iterations produce even more accurate estimates. ☒

Approximate linearization also works well for some rapidly growing sequences.

Example 9.8 (*Doubly exponential sequences*). Many recurrences are of the form

$$a_{n+1} = a_n^2 + b_n , \tag{9.56}$$

where b_n is much smaller than a_n^2 (and may even depend on the a_n for $k \leqslant n$, as in $b_n = a_n$ or $b_n = a_{n-1}$). Aho and Sloane (1973) found that surprisingly simple solutions to such recurrences can often be found. The basic idea is to reduce to approximate linearization by taking logarithms. We find that if a_0 is the given initial value, and $a_n > 0$ for all n, then the transformation

$$u_n = \log a_n , \tag{9.57}$$

$$\delta_n = \log(1 + b_n a_n^{-2}) , \tag{9.58}$$

reduces (9.56) to

$$u_{n+1} = 2u_n + \delta_n , \quad n \geqslant 0 . \tag{9.59}$$

Therefore

$$\begin{aligned}u_n &= \delta_{n-1} + 2u_{n-1} = \delta_{n-1} + 2\delta_{n-2} + 4u_{n-2} \\ &= \ldots \\ &= \sum_{j=1}^{n} 2^{j-1}\delta_{n-j} + 2^n u_0 \\ &= 2^n(u_0 + \delta_0/2 + \delta_1/4 + \cdots + \delta_{n-1}/2^n) .\end{aligned} \tag{9.60}$$

If we assume that the δ_k are small, then

$$\alpha = u_0 + \sum_{k=0}^{\infty} \delta_k 2^{-k-1} \tag{9.61}$$

exists, and

$$r_n = u_n - 2^n\alpha = 2^n \sum_{k=n}^{\infty} \delta_k 2^{-k-1} \,. \tag{9.62}$$

If the δ_k are sufficiently small, the difference r_n in (9.62) will be small, and

$$a_n = \exp(u_n) = \exp(2^n\alpha - r_n) \,. \tag{9.63}$$

The expression (9.63) might not seem satisfactory, since both a_n and r_n are expressed in terms of all the a_k, for $k < n$ and for $k \geqslant n$. The point of (9.63) is that for many recurrences, r_n is negligibly small, while α is given by the rapidly convergent series (9.61), so that only the first few a_n are needed to obtain a good estimate for the asymptotic behavior of a_n. We next discuss a particularly elegant case.

Suppose that $a_n \geqslant 1$ and $|b_n| < a_n/4$ for all $n \geqslant 0$. Then $a_{n+1} \geqslant a_n$ and $|\delta_{n+1}| \leqslant |\delta_n|$ for $n \geqslant 0$, and so $|r_n| \leqslant |\delta_n|$. Hence

$$a_n \exp(-|\delta_n|) \leqslant \exp(2^n\alpha) \leqslant a_n \exp(|\delta_n|) \tag{9.64}$$

and since

$$\begin{aligned} \exp(|\delta_n|) &\leqslant 1 + |b_n| a_n^{-2} \ < \ 1 + (4a_n)^{-1} \,, \\ \exp(-|\delta_n|) &\geqslant (1 + (4a_n)^{-1})^{-1} \ \geqslant \ 1 - (3a_n)^{-1} \,, \end{aligned} \tag{9.65}$$

we find that

$$|a_n - \exp(2^n\alpha)| < (2a_n)^{-1} \leqslant 1/2 \,. \tag{9.66}$$

If a_n is an integer, then we can assert that it is the closest integer to $\exp(2^n\alpha)$.

The restriction $|b_n| < a_n/4$ is severe. The basic method applies even without it, and the expansion (9.63) is valid, for example, if we only require that $|\delta_{n+1}| \leqslant |\delta_n|$ for $n \geqslant n_0$. However, we will not in general obtain results as nice as (9.66) if we only impose these weak conditions.

The method outlined above can be applied to recurrences that appear to be of a slightly different form. Sometimes only a trivial transformation is required. For example, Golomb's nonlinear recurrence,

$$a_{n+1} = a_0 a_1 \cdots a_n + b, \quad a_0 = 1 \,, \tag{9.67}$$

for b a constant, is easily seen to be equivalent to

$$a_{n+1} = (a_n - b)a_n + b, \quad a_0 = 1, \quad a_1 = b + 1 \,. \tag{9.68}$$

The substitution

$$x_n = a_n - b/2 \tag{9.69}$$

transforms (9.68) into

$$x_{n+1} = x_n^2 + (2-b)b/4 , \tag{9.70}$$

which is of the form treated above. [If the x_n are integers, the inequality (9.66) with x_n replacing a_n might not apply to the x_n because the condition $|(2-b)b/4| < |x_k|/4$ might fail for some k. The trick to use here is to start the recurrence with some x_k, say x_{k_0}, so that the condition $|(2-b)b/4| < |x_k|/4$ applies for $k \geqslant k_0$. The new α for which (9.66) holds will then be defined in terms of $x_{k_0}, x_{k_0+1}, \ldots$.]

In some situations the results presented above cannot be applied, but the basic method can still be extended. That is the case for the recurrence

$$a_{n+1} = a_n a_{n-1} + 1, \quad a_0,\ a_1 \geqslant 1 \tag{9.71}$$

of Aho and Sloane (1973). The result is that a_n is the nearest integer to

$$\alpha^{F_n} \beta^{F_{n-1}} , \tag{9.72}$$

where α and β are positive constants, and the F_k are the Fibonacci numbers. What matters is that the recurrence leads to doubly exponential (and regular) growth of a_n. Example 15.3 shows how this principle can be applied even when the a_n are not numbers, but polynomials whose coefficients need to be estimated. ☒

9.5. *Quasi-linear recurrences*

This section mentions some methods and results for studying recurrences that have linearity properties, but are not linear. The most important of them are recurrences involving minimization or maximization. They arise frequently in problems that use dynamic programming approaches and in divide-and-conquer methods. An important example, treated in Fredman and Knuth (1974), is that of a sequence f_n, given by $f_0 = 1$ and

$$f_{n+1} = g_{n+1} + \min_{0 \leqslant k \leqslant n} (\alpha f_k + \beta f_{n-k}) \quad \text{for } n \geqslant 0 , \tag{9.73}$$

where $\alpha, \beta > 0$, and g_n is some given sequence. Fredman and Knuth showed that if $g_n = 0$ for $n \geqslant 1$ and $\alpha + \beta < 1$, then

$$f_n \geqslant c n^{1+1/\gamma} \quad \text{for some } c = c(\alpha, \beta) > 0 , \tag{9.74}$$

where γ is the solution to

$$\alpha^{-\gamma} + \beta^{-\gamma} = 1 . \tag{9.75}$$

They proved that $\lim_{n\to\infty} f_n n^{-1-1/\gamma}$ exists if and only if $(\log \alpha)/(\log \beta)$ is irrational. They also presented analyses of this recurrence for $\alpha + \beta \geqslant 1$, as well as of several recurrences that have different g_n.

The value of the Fredman–Knuth paper is less in the precise results they obtain for several recurrences of the type (9.73) than in the methods they develop, which allow one to analyze related problems. A crucial role in their approach is played by the observation that for the g_n they consider, the minimum in (9.73) can be located rather precisely. The conditions for such localization are applicable to many other sequences as well.

Further work on the recurrence (9.73) was done by Kapoor and Reingold (1985), who obtained a complete solution under certain conditions. Their solution is complicated, expressed in terms of the weighted external path length of a binary tree. It is sufficiently explicit, though, to give a complete picture of the continuity, convexity, and oscillation properties of f_n. In some cases their solution simplifies dramatically.

Another class of quasi-linear recurrences involves the greatest-integer function. Following Erdős et al. (1987), consider recurrences of the form

$$a(0) = 1, \qquad a(n) = \sum_{i=1}^{s} r_i a(\lfloor n/m_i \rfloor), \quad n \geqslant 1, \tag{9.76}$$

where $r_i > 0$ for all i, and the m_i are integers, $m_i \geqslant 2$ for all i. Let $\tau > 0$ be the (unique) solution to

$$\sum_{i=1}^{s} r_i m_i^{-\tau} = 1 \ . \tag{9.77}$$

If there is an integer d and integers u_i such that $m_i = d^{u_i}$ for $1 \leqslant i \leqslant s$, then lim $a(n)n^{-\tau}$ as $n \to \infty$ i does not exist, but the limit of $a(d^k)d^{-k\tau}$ as $k \to \infty$ does exist. If there is no such d, then the limit of $a(n)n^{-\tau}$ as $n \to \infty$ does exist, and can readily be computed. For example, when

$$a(n) = a(\lfloor n/2 \rfloor) + a(\lfloor n/3 \rfloor) + a(\lfloor n/6 \rfloor) \quad \text{for } n \geqslant 1 \ ,$$

this limit is $12(\log 432)^{-1}$. Convergence to the limit is extremely slow, as is shown in Erdős et al. (1987). The method of proof used in that paper is based on renewal theory. Several other methods for dealing with recurrences of the type (9.76) are mentioned there and in the references listed in that paper. There are connections to other recurrences that are linear in two variables, such as

$$b(m,n) = b(m,n-1) + b(m-1,n) + b(m-1,n-1), \quad m,n \geqslant 1 \ . \tag{9.78}$$

Consider an infinite sequence of integers $2 \leqslant a_1 < a_2 < \cdots$ such that

$$\sum_{j=1}^{\infty} a_j^{-1} \log a_j < \infty \ ,$$

and define $c(0) = 0$,

$$c(n) = \sum_{j=1}^{\infty} c(\lfloor n/a_j \rfloor) + 1, \quad n \geqslant 1 . \tag{9.79}$$

If ρ is the (unique) positive solution to

$$\sum_{j=1}^{\infty} a_j^{-\rho} = 1 ,$$

then Erdős (1941) showed that

$$c(n) \sim cn^{\rho} \quad \text{as } n \to \infty \tag{9.80}$$

for a positive constant c. Although the recurrence (9.79) is similar to that of eq. (9.76), the results are different [no oscillations can occur for a recurrence given by eq. (9.79)] and the methods are dissimilar.

Karp (1991) considers recurrences of the type $T(x) = a(x) + T(h(x))$, where x is a nonnegative real variable, $a(x) \geqslant 0$, and $h(x)$ is a random variable, $0 \leqslant h(x) \leqslant x$, with $m(x)$ being the expectation of $h(x)$. Such recurrences arise frequently in the analysis of algorithms, and Karp proves several theorems that bound the probability that $T(x)$ is large. For example, he obtains the following result.

Theorem 9.9. *Suppose that $a(x)$ is a nondecreasing continuous function that is strictly increasing on $\{x: a(x) > 0\}$, and $m(x)$ is a continuous function. Then for all $x \in \mathbb{R}^+$ and $k \in \mathbb{Z}^+$,*

$$\operatorname{Prob}(T(x) \geqslant u(x) + ka(x)) \leqslant (m(x)/x)^k ,$$

where $u(x)$ is the unique least nonnegative solution to the equation $u(x) = a(x) + u(m(x))$.

Another result, proved in Greenberg et al. (1988), is the following estimate.

Theorem 9.10. *Suppose that $r, a_1, \ldots, a_N \in \mathbb{R}^+$ and that $b \geqslant 0$. For $n > N$, define*

$$a_n = 1 + \max_{1 \leqslant k \leqslant n-1} \frac{b + a_{n-1} + a_{n-2} + \cdots + a_{n-k}}{k + r} . \tag{9.81}$$

Then

$$a_n \sim (n/r)^{1/2} \quad \textit{as } n \to \infty . \tag{9.82}$$

Theorem 9.10 is proved by an involved induction on the behavior of the a_n.

10. Analytic generating functions

Combinatorialists use recurrence, generating functions, and such transformations as the Vandermonde convolution; others, to my horror, use *contour integrals, differential equations, and other resources of mathematical analysis.*

J. Riordan (1968)

The use of analytic methods in combinatorics did horrify Riordan. They are widespread, though, because of their utility, which even Riordan could not deny. About half of this chapter is devoted to such methods, as they are extremely flexible and give very precise estimates.

10.1. Introduction and general estimates

This section serves as an introduction to most of the remaining sections of the paper, which are concerned largely with the use of methods of complex variables in asymptotics. Many of the results to be presented later can be used with little or no knowledge of analytic functions. However, even some slight knowledge of complex analysis is helpful in getting an understanding of the scope and limitations of the methods to be discussed. There are many textbooks on analytic functions, such as Henrici (1974–86) and Titchmarsh (1939). This chapter assumes that the reader has some knowledge of this field, but not a deep one. It reviews the concepts that are most relevant in asymptotic enumeration, and how they affect the estimates that can be obtained. It is not a general introduction to the subject of complex analysis, and the choices of topics, their ordering, and the decision of when to include proofs were all made with the goal of illustrating how to use complex analysis in asymptotics.

There are several definitions of analytic functions, all equivalent. For our purposes, it will be most convenient to call a function $f(z)$ of one complex variable *analytic* in a connected open set $S \subseteq \mathbb{C}$ if in a small neighborhood of every point $w \in S$, $f(z)$ has an expansion as a power series

$$f(z) = \sum_{n=0}^{\infty} a_n(z-w)^n, \quad a_n = a_n(w), \tag{10.1}$$

that converges. Practically all the functions encountered in asymptotic enumeration that are analytic are analytic in a disk about the origin. A necessary and sufficient condition for $f(z)$, defined by a power series (6.1), to be analytic in a neighborhood of the origin is that $|a_n| \leqslant C^n$ for some constant $C > 0$. Therefore there is an effective dichotomy, with common generating functions either not converging near 0 and being only formal power series, or else converging and being analytic.

A function $f(z)$ is called *meromorphic* in S if it is analytic in S except at a (countable isolated) subset $S' \subseteq S$, and in a small neighborhood of every $w \in S'$, $f(z)$ has an expansion of the form

$$f(z) = \sum_{n=-N(w)}^{\infty} a_n(z-w)^n , \quad a_n = a_n(w) . \tag{10.2}$$

Thus meromorphic functions can have poles, but nothing more. Alternatively, a function is meromorphic in S if and only if it is the quotient of two functions analytic in S. In particular, z^{-5} is meromorphic throughout the complex plane, but $\sin(1/z)$ is not. In general, functions given by nice expressions are analytic away from obvious pathological points, since addition, multiplication, division, and composition of analytic functions usually yield analytic or meromorphic functions in the proper domains. Thus $\sin(1/z)$ is analytic throughout $\mathbb{C} \setminus \{0\}$, and so is z^{-5}, while $\exp(1/(1-z))$ is analytic throughout $\mathbb{C} \setminus \{1\}$, but is not meromorphic because of the essential singularity at $z = 1$. Not all functions that might seem smooth are analytic, though, as neither $f(z) = \bar{z}$ ($\bar{z}$ denoting the complex conjugate of z) nor $f(z) = |z|$ is analytic anywhere. The smoothness condition imposed by (10.1) is very stringent.

Analytic continuation is an important concept. A function $f(z)$ may be defined and analytic in S, but there may be another function $g(z)$ that is analytic in $S' \supset S$ and such that $g(z) = f(z)$ for $z \in S$. In that case we say that $g(z)$ provides an analytic continuation of $f(z)$ to S', and it is a theorem that this extension is unique. A simple example is provided by

$$\sum_{n=0}^{\infty} z^n = \frac{1}{1-z} \,. \tag{10.3}$$

The power series on the left side converges only for $|z| < 1$, and defines an analytic function there. On the other hand, $(1-z)^{-1}$ is analytic throughout $\mathbb{C} \setminus \{1\}$, and so provides an analytic continuation for the power series. This is a common phenomenon in asymptotic enumeration. Typically a generating function will converge in a disk $|z| < r$, will have a singularity at r, but will be continuable to a region of the form

$$\{z\colon |z| < r + \delta,\ |\mathrm{Arg}(z - r)| > \pi/2 - \varepsilon\} \tag{10.4}$$

for δ, $\varepsilon > 0$. When this happens, it can be exploited to provide better or easier estimates of the coefficients, as is shown in section 11.1. That section explains the reasons why continuation to a region of the form (10.4) is so useful.

If $f(z)$ is analytic in S, z is on the boundary of S, but $f(z)$ cannot be analytically continued to a neighborhood of z, we say that z is a *singularity* of $f(z)$. Isolated singularities that are not poles are called essential, so that $z = 1$ is an essential singularity of $\exp(1/(1-z))$, but not of $1/(1-z)$. (Note that $z = 1$ is an essential singularity of $f(z) = (1-z)^{1/2}$ even though $f(1) = 0$.) Throughout the rest of this chapter we will often refer to *large singularities* and *small singularities*. These are not precise concepts, and are meant only to indicate how fast the function $f(z)$ grows as $z \to z_0$, where z_0 is a singularity. If $z_0 = 1$, we say that $(1-z)^{1/2}$, $\log(1-z)$, $(1-z)^{-10}$ have small singularities, since $|f(z)|$ either decreases or grows at most like a negative power of $|1-z|$ as $z \to 1$. On the other hand, $\exp(1/(1-z))$ or $\exp((1-z)^{-1/5})$ will be said to have large singularities. Note that for $z = 1 + iy$, $y \in \mathbb{R}$, $\exp(1/(1-z))$ is bounded, so the choice of path along which the singularity is approached is important. In determining the size of a singularity z_0, we will usually be concerned with real z_0 and generating functions $f(z)$ with nonnegative coefficients, and then usually will need to look only at z real, $z \to z_0^-$. When the function $f(z)$ is *entire* (that is, analytic throughout $\mathbb{C}$), we will say that ∞ is a singularity of $f(z)$ (unless $f(z)$ is a constant), and will use the large vs. small singularity classification depending on how fast $f(z)$ grows as $|z| \to \infty$. The distinction between small and large singularities is important in asymptotics because different methods are used in the two cases.

A simple closed contour Γ in the complex plane is given by a continuous mapping $\gamma : [0,1] \to \mathbb{C}$ with the properties that $\gamma(0) = \gamma(1)$, and that $\gamma(s) \neq \gamma(t)$ whenever $0 \leqslant s < t \leqslant 1$ and either $s \neq 0$ or $t \neq 1$. Intuitively, Γ is a closed path in the complex plane that does not intersect itself. For most applications that will be made in this chapter, simple closed contours Γ will consist of line segments and sections of circles. For such contours it is easy to prove that the complex plane is divided by the contour into two connected components, the inside and the outside of the curve. This result is true for all simple closed curves by the Jordan curve theorem, but this result is surprisingly hard to prove.

In asymptotic enumeration, the basic result about analytic functions is the Cauchy integral formula for their coefficients.

Theorem 10.1. *If $f(z)$ is analytic in an open set S containing* 0, *and*

$$f(z) = \sum_{n=0}^{\infty} a_n z^n \tag{10.5}$$

in a neighborhood of 0, *then for any $n \geqslant 0$,*

$$a_n = [z^n]f(z) = (2\pi i)^{-1} \int_\Gamma f(z) z^{-n-1}\, dz\ , \tag{10.6}$$

where Γ is any simple closed contour in S that contains the origin inside it and is positively oriented (i.e., traversed in counterclockwise direction).

An obvious question is why should one use the integral formula (10.6) to determine the coefficient a_n of $f(z)$. After all, the series (10.5) shows that

$$n!\, a_n = \frac{d^n}{dz^n} f(z)\Big|_{z=0}\ . \tag{10.7}$$

Unfortunately the differentiation involved in (10.7) is hard to control. Derivatives involve taking limits, and so even small changes in a function can produce huge changes in derivatives, especially high-order ones. The special properties of analytic functions are not reflected in the formula (10.7), and for nonanalytic functions there is little that can be done. On the other hand, Cauchy's integral formula (10.6) does use special properties of analytic functions, which allow the determination of the coefficients of $f(z)$ from the values of $f(z)$ along any closed path. This determination involves integration, so that even coarse information about the size of $f(z)$ can be used with it. The analytic methods that will be outlined exploit the freedom of choice of the contour of integration to relate the behavior of the coefficients to the behavior of the function near just one or sometimes a few points.

If the power series (10.5) converges for $|z| < R$, and for the contour Γ we choose a circle $z = r\exp(i\theta)$, $0 \leqslant \theta \leqslant 2\pi$, $0 < r < R$, then the validity of (10.6) is easily checked by direct computation, since the power series converges absolutely and uniformly so one can interchange integration and summation. The strength of Cauchy's formula is in the freedom to choose the contour Γ in different ways. This freedom yields most of the powerful results to be discussed in the following sections, and later in this section we will outline how this is achieved. First we discuss some simple applications of Theorem 10.1 obtained by choosing Γ to be a circle centered at the origin.

Theorem 10.2. *If $f(z)$ is analytic in $|z| < R$, then for any r with $0 < r < R$ and any $n \in \mathbb{Z}$, $n \geqslant 0$,*

$$|[z^n]f(z)| \leqslant r^{-n} \max_{|z|=r} |f(z)| \, . \tag{10.8}$$

The choice of Γ in Theorem 10.1 to be the circle of radius r gives Theorem 10.2. If $f(z)$, defined by (10.5), has $a_n \geqslant 0$ for all n, then

$$|f(z)| \leqslant \sum_{n=0}^{\infty} a_n |z|^n = f(|z|)$$

and therefore we obtain Lemma 8.1 as an easy corollary to Theorem 10.2. The advantage of Theorem 10.2 over Lemma 8.1 is that there is no requirement that $a_n \geqslant 0$. The bound of Theorem 10.2 is usually weaker than the correct value by a small multiplicative factor such as $n^{1/2}$.

If $f(z)$ is analytic in $|z| < R$, then for any $\delta > 0$, $f(z)$ is bounded in $|z| < R - \delta$, and so Theorem 10.2 shows that $a_n = [z^n]f(z)$ satisfies $|a_n| = \mathrm{O}((R - \delta)^{-n})$. On the other hand, if $|a_n| = \mathrm{O}(S^{-n})$, then the power series (10.5) converges for $|z| < S$ and defines an analytic function in that disk. Thus we obtain the easy result that if $f(z)$ is analytic in a disk $|z| < R$ but in no larger disk, then

$$\limsup |a_n|^{1/n} = R^{-1} \, . \tag{10.9}$$

Example 10.3 (*Oscillating sequence*). Consider the sequence, discussed already in Example 9.5, given by

$$a_n = \sum_{k=0}^{n} \binom{n}{k} \frac{(-1)^k}{k!} \, , \qquad n = 0, 1, \ldots \, . \tag{10.10}$$

The maximal term in the sum (10.10) is of order roughly $\exp(cn^{1/2})$, so a_n cannot be much larger. However, the sum (10.10) does not show that a_n cannot be extremely small. Could we have $|a_n| \leqslant \exp(-n)$ for all n, say? That this is impossible is obvious from (9.39), though, by the argument above. The generating function $f(z)$, given by eq. (9.39), is analytic in $|z| < 1$, but has an essential singularity at

$z = 1$, so we immediately see that for any $\varepsilon > 0$, $|a_n| < (1+\varepsilon)^n$ for all sufficiently large n, and that $|a_n| > (1-\varepsilon)^n$ for infinitely many n. (More powerful methods for dealing with analytic generating functions, such as the saddle point method to be discussed in section 12, can be used to obtain the asymptotic relation for a_n given in Example 9.5.) ⊠

There is substantial literature dealing with the growth rate of coefficients of analytic functions. The book of Evgrafov (1961) is a good reference for these results. However, the estimates presented there are not too useful for us, since they apply to wide classes of often pathological functions. In combinatorial enumeration we usually encounter much tamer generating functions for which the crude bounds of Evgrafov (1961) are obvious or easy to derive. Instead, we need to use the tractable nature of the functions we encounter to obtain much more delicate estimates.

The basic result, derived earlier, is that the power-series coefficients a_n of a generating function $f(z)$, defined by (10.5), grow in absolute value roughly like R^{-n}, if $f(z)$ is analytic in $|z| < R$. A basic result about analytic functions says that if the Taylor series (10.5) of $f(z)$ converges for $|z| < R$ but for every $\varepsilon > 0$ there is a z with $|z| = R + \varepsilon$ such that the series (10.5) diverges at z, then $f(z)$ has a singularity z with $|z| = R$. Thus the exponential growth rate of the a_n is determined by the distance from the origin of the nearest singularity of $f(z)$, with close singularities giving large coefficients. Sometimes it is not obvious what R is. When the coefficients of $f(z)$ are positive, as is common in combinatorial enumeration and analysis of algorithms, there is a useful theorem of Pringsheim (Titchmarsh 1939):

Theorem 10.4. *Suppose that $f(z)$ is defined by eq.* (10.5) *with $a_n \geqslant 0$ for all $n \geqslant n_0$, and that the series* (10.5) *for $f(z)$ converges for $|z| < R$ but not for any $|z| > R$. Then $z = R$ is a singularity of $f(z)$.*

As we remarked above, the exponential growth rate of the a_n is determined by the distance from the origin of the nearest singularity. Theorem 10.4 says that if the coefficients a_n are nonnegative, it suffices to look along the positive real axis to determine the radius of convergence R, which is also the desired distance to the singularity. There can be other singularities at the same distance from the origin (for example, $f(z) = (1-z^2)^{-1}$ has singularities at $z = \pm 1$), but Theorem 10.4 guarantees that none are closer to 0 than the positive real one.

Since the singularities of smallest absolute value of a generating function exert the dominant influence on the asymptotics of the corresponding sequence, they are called the *dominant singularities*. In the most common case there is just one dominant singularity, and it is almost always real. However, we will sometimes speak of a large set of singularities (such as the k first-order poles in Theorem 9.2, which are at different distances from the origin) as dominant ones. This allows some dominant singularities to be more influential than others.

Many techniques, including the elementary methods of section 8, obtain bounds for summatory functions of coefficients even when they cannot estimate the individual coefficients. These methods succeed largely because they create a dominant singularity. If $f(z) = \sum f_n z^n$ converges for $|z| < 1$, diverges for $|z| > 1$, and has $f_n \geqslant 0$, then the singularity at $z = 1$ is at least as large as any other. However, there could be other singularities on $|z| = 1$ that are just as large. [This holds for the functions $f_2(z)$ and $f_3(z)$ defined by (8.2) and (8.4).] When we consider the generating function of $\sum_{k \leqslant n} f_k$, though, we find that

$$h(z) = \sum_{n=0}^{\infty} \left(\sum_{k=0}^{n} f_k \right) z^n = (1 - z)^{-1} f(z) , \tag{10.11}$$

so that $h(z)$ has a singularity at $z = 1$ that is much larger than any other one. That often provides enough of an extra boost to push through the necessary technical details of the estimates.

Most generating functions $f(z)$ have their coefficients $a_n = [z^n]f(z)$ real. If $f(z)$ is analytic at 0, and has real coefficients, then $f(z)$ satisfies the reflection principle,

$$f(z) = \overline{f(\overline{z})} . \tag{10.12}$$

This implies that zeros and singularities of $f(z)$ come in complex-conjugate pairs.

The success of analytic methods in asymptotics comes largely from the use of Cauchy's formula (10.6) to estimate accurately the coefficients a_n. At a more basic level, this success comes because the behavior of an analytic function $f(z)$ reflects precisely the behavior of the coefficients a_n. In the discussion of elementary methods in section 8, we pointed out that the behavior of a generating function for real arguments does not distinguish between functions with different coefficients. For example, the functions $f_1(z)$ and $f_3(z)$ defined by (8.1) and (8.4) are almost indistinguishable for $z \in \mathbb{R}$. However, they differ substantially in their behavior for complex z. The function $f_1(z)$ has only a first-order pole at $z = 1$ and no other singularities, while $f_3(z)$ has poles at $z = 1$, $\exp(2\pi i/3)$, and $\exp(4\pi i/3)$. The three poles at the three cubic roots of unity reflect the modulo 3 periodicity of the coefficients of $f_3(z)$. This is a general phenomenon, and in the next section we sketch the general principle that underlies it. (The degree to which coefficients of an analytic function determine the behavior at the singularities is the subject of Abelian theorems. We will not need to delve into this subject to its full depth. For references, see Hardy (1949) and Titchmarsh (1939).]

Analytic methods are extremely powerful, and when they apply, they often yield estimates of unparalleled precision. However, there are tricky situations where analytic methods seem as if they ought to apply, but do not (at least not easily), whereas simpler approaches work.

Example 10.5 (*Set partitions with distinct block sizes*). Let a_n be the number of partitions of a set of n elements into blocks of distinct sizes. Then $a_n = b_n \cdot n!$, where $b_n = [z^n]f(z)$, with

$$f(z) = \prod_{k=1}^{\infty} \left(1 + \frac{z^k}{k!} \right) . \tag{10.13}$$

The function $f(z)$ is entire and has nonnegative coefficients, so it might appear as an ideal candidate for an application of some of the methods for dealing with large singularities (such as the saddle point technique) that will be presented later. However, on circles $|z| = (n + 1/2)/\mathrm{e}$, $n \in \mathbb{Z}^+$, $f(z)$ does not vary much, so there are technical problems in applying these analytic methods. On the other hand, combinatorial estimates can be used to show (Knopfmacher et al. 1995) that the b_n behave in a "regularly irregular" way, so that, for example,

$$b_{m(m+1)/2-1} \sim b_{m(m+1)/2} \quad \text{as } m \to \infty , \tag{10.14}$$

$$b_{m(m+1)/2} \sim m b_{m(m+1)/2+1} \quad \text{as } m \to \infty . \tag{10.15}$$

These estimates are obtained by expanding the product in eq. (10.13) and noting that

$$b_n = \sum_{\substack{r, 1 \leqslant k_1 < \cdots < k_r \\ \sum k_i = n}} \frac{1}{\prod_{i=1}^{r} k_i!} . \tag{10.16}$$

Since factorials grow rapidly, the only terms in the sum in (10.16) that are significant are those with small k_i. The term $b_n z^n$ for $n = m(m+1)/2$ for example, comes almost entirely from the product of $z^k/k!$, $1 \leqslant k \leqslant m$, all other products contributing an asymptotically negligible amount. ⊠

10.2. Subtraction of singularities

An important basic tool in asymptotics of coefficients of analytic functions is that of subtraction of singularities. If we wish to estimate $[z^n]f(z)$, and we know $[z^n]g(z)$, and the singularities of $f(z) - g(z)$ are smaller than those of $f(z)$, then we can usually conclude that $[z^n]f(z) \sim [z^n]g(z)$ as $n \to \infty$. In practice, given a function $f(z)$, we find the dominant singularities of $f(z)$ (usually poles), and construct a simple function $g(z)$ with those singularities. We illustrate this approach with several examples. The basic theme will recur in other sections.

Example 10.6 (*Bernoulli numbers*). The Euler–Maclaurin summation formula, introduced in section 5.3, involves the Bernoulli numbers B_n with exponential generating function

$$f(z) = \sum_{n=0}^{\infty} B_n \frac{z^n}{n!} = \frac{z}{\mathrm{e}^z - 1} . \tag{10.17}$$

The denominator $\exp(z) - 1$ has zeros at 0, $\pm 2\pi\mathrm{i}$, $\pm 4\pi\mathrm{i}, \ldots$. The zero at 0 is canceled by the zero of z, so $f(z)$ is analytic for $|z| < 2\pi$, but has first-order poles at $z = \pm 2\pi\mathrm{i}$, $\pm 4\pi\mathrm{i}, \ldots$. Consider

$$g(z) = 2\pi\mathrm{i} \left(\frac{1}{z - 2\pi\mathrm{i}} - \frac{1}{z + 2\pi\mathrm{i}} \right) . \tag{10.18}$$

Then $f(z) - g(z)$ is analytic for $|z| < 4\pi$, so

$$|[z^n](f(z) - g(z))| = O((4\pi - \varepsilon)^{-n}) \quad \text{as } n \to \infty \tag{10.19}$$

for every $\varepsilon > 0$. On the other hand,

$$[z^n]g(z) = \begin{cases} 0 & n \text{ odd}, \\ 2(2\pi)^{-n} & n \text{ even}. \end{cases} \tag{10.20}$$

This gives the leading term asymptotics of B_n. By taking more complicated $g(z)$, we can subtract more of the singularities of $f(z)$ and obtain more accurate expansions for B_n. It is even possible to obtain an exponentially rapidly convergent series for B_n. ⊠

Example 10.7 (*Rational function asymptotics*). As another example of the subtraction-of-singularities principle, we sketch a proof of Theorem 9.2. Suppose that the hypotheses of that theorem are satisfied. Let

$$u(z) = \sum_{j=1}^{k} \frac{-g(\rho_j)}{\rho_j h'(\rho_j)(1 - z/\rho_j)} \,. \tag{10.21}$$

Then $f(z) - u(z)$ has no singularities in $|z| \leqslant R$, and for $|z| = R$,

$$|f(z) - u(z)| \leqslant |f(z)| + |u(z)| \leqslant W + \delta^{-1} \sum_{j=1}^{k} |g(\rho_j)/h'(\rho_j)| \,. \tag{10.22}$$

Hence, by Theorem 10.2,

$$\left|[z^n](f(z) - u(z))\right| \leqslant WR^{-n} + \delta^{-1}R^{-n} \sum_{j=1}^{k} |g(\rho_j)/h'(\rho_j)| \,. \tag{10.23}$$

On the other hand,

$$[z^n]u(z) = -\sum_{j=1}^{k} \rho_j^{-n-1} g(\rho_j)/h'(\rho_j) \,. \tag{10.24}$$

The last two estimates yield Theorem 9.2. ⊠

The reader may have noticed that the proof of Theorem 9.2 presented above does not depend on $f(z)$ being rational. We have proved the following more general result.

Theorem 10.8. *Suppose that $f(z)$ is meromorphic in an open set containing $|z| \leqslant R$, that it is analytic at $z = 0$ and on $|z| = R$, and that the only poles of $f(z)$ in $|z| < R$ are at $\rho_1, \ldots, \rho_k$, each of multiplicity* 1. *Suppose further that*

$$\max_{|z|=R} |f(z)| \leqslant W \tag{10.25}$$

and that $R - |\rho_j| \geqslant \delta$ for some $\delta > 0$ and $1 \leqslant j \leqslant k$. Then

$$\left| [z^n]f(z) + \sum_{j=1}^{k} r_j \rho_j^{-n-1} \right| \leqslant WR^{-n} + \delta^{-1}R^{-n}\sum_{j=1}^{k} |r_j| \,, \tag{10.26}$$

where r_j is the residue of $f(z)$ at ρ_j.

In the examples above, the dominant singularities were separated from other ones, so their contributions were larger than those of lower-order terms by an exponential factor. Sometimes the singularity that remains after subtraction of the dominant one is on the same circle, and only slightly smaller. Section 11 presents methods that deal with some cases of this type, at least when the singularity is not large. What makes those methods work is the subtraction-of-singularities principle. Next we illustrate another application of this principle where the singularity is large. (The generating function is entire, and so the singularity is at infinity.)

Example 10.9 (*Permutations without long increasing subsequences*). Let $u_k(n)$ be the number of permutations of $\{1,2,\ldots,n\}$ that have no increasing subsequence of length $> k$. Logan and Shepp (1977) and Vershik and Kerov (1977) established by calculus of variations and combinatorics that the average value of the longest increasing subsequence in a random permutation is asymptotic to $2n^{1/2}$. Frieze (1991) has proved recently, using probabilistic methods, a stronger result, namely that almost all permutations have longest increasing subsequences of length close to $2n^{1/2}$. Here we consider asymptotics of $u_k(n)$ for k fixed and $n \to \infty$. The Schensted correspondence and the hook formula express $u_k(n)$ in terms of Young diagrams with $\leqslant k$ columns. For k fixed, there are few diagrams and their influence can be estimated explicitly using Stirling's formula, although Selberg-type integrals are involved and the analysis is complicated. This analysis was done by Regev (1981), who proved more general results. Here we sketch another approach to the asymptotics of $u_k(n)$ for k fixed. It is based on a result of Gessel (1990). If

$$U_k(z) = \sum_{n=0}^{\infty} \frac{u_k(n)z^{2n}}{(n!)^2} \,, \tag{10.27}$$

then

$$U_k(z) = \det(I_{|i-j|}(2z))_{1\leqslant i,j\leqslant k} \,, \tag{10.28}$$

where the $I_m(x)$ are Bessel functions (chapter 9 of NBS 1970). H. Wilf and the author have noted that one can obtain the asymptotics of the $u_k(n)$ by using known asymptotic results about the $I_m(x)$. Equation (9.7.1) of NBS 1970 states that for every $H \in \mathbb{Z}^+$,

$$I_m(z) = (2\pi z)^{-1/2} e^z \left(\sum_{h=0}^{H-1} c(m,h) z^{-h} + O(|z|^{-H}) \right) \,, \tag{10.29}$$

where this expansion is valid for $|z| \to \infty$ with $|\mathrm{Arg}(z)| \leqslant 3\pi/8$, say. The $c(m,h)$ are explicit constants with $c(m,0) = 1$. Let us consider $k = 4$ to be concrete. Then, taking $H = 7$ in (10.29) (higher values of H are needed for larger k) we find from (10.28) that

$$U_4(z) = \mathrm{e}^{8z}(3(256\pi^2 z^8)^{-1} + \mathrm{O}(|z|^{-9})) \quad \text{for } |z| \geqslant 1 . \tag{10.30}$$

It is also known that $I_m(-z) = (-1)^m I_m(z)$ and $I_m(z)$ is relatively small in the angular region $|\pi/2 - \mathrm{Arg}(z)| < \pi/8$. Therefore $U_4(-z) = U_4(z)$, and one can show that

$$|U_4(z)| = \mathrm{O}(|z|^{-1} U_4(|z|)) \tag{10.31}$$

for z away from the real axis.

To apply the subtraction-of-singularities principle, we need an entire function $f(z)$ that is even, is large only near the real axis, and such that for $x \in \mathbb{R}$, $x \to \infty$,

$$f(x) \sim 3(256\pi^2 x^8)^{-1} \exp(8x) . \tag{10.32}$$

The function

$$f^*(z) = 3(128\pi^2 z^8)^{-1} \cosh(8z)$$

is even and has the desired asymptotic growth, but is not entire. We correct this defect by subtracting the contribution of the pole at $z = 0$, and let

$$\begin{aligned} f(z) = 3(128\pi^2 z^8)^{-1}(\cosh(8z) - 1 - 32z^2 - 512z^4/3 - 16\,384z^6/45 \\ - 131\,072z^8/315) . \end{aligned} \tag{10.33}$$

(It is not necessary to know explicitly the first eight terms in the Taylor expansion of $\cosh(8z)$ that we wrote down above, as they do not affect the final answer.) With this definition

$$|U_4(z) - f(z)| = \mathrm{O}(|z|^{-1} f(|z|)) \tag{10.34}$$

uniformly for all z with $|z| \geqslant 1$, say, and so if we apply Cauchy's theorem on the circle $|z| = n/4$, say, we find that

$$[z^{2n}](U_4(z) - f(z)) = \mathrm{O}(n^{-2n} \mathrm{e}^{2n} 16^n n^{-9}) . \tag{10.35}$$

(The choice of $|z| = n/4$ is made to minimize the resulting estimate.) On the other hand, by Stirling's formula,

$$\begin{aligned} [z^{2n}]f(z) &= 3(128\pi^2)^{-1} \cdot ([z^{2n+8}]\cosh(8z)) \\ &= 3(128\pi^2)^{-1} 8^{2n+8}/(2n+8)! \\ &\sim 1536\pi^{-5/2} n^{-2n} 16^n \mathrm{e}^{2n} n^{-17/2} \quad \text{as } n \to \infty . \end{aligned} \tag{10.36}$$

Comparing (10.35) and (10.36), we see that

$$\begin{aligned} u_4(n) = (n!)^2 [z^{2n}] U_4(z) \sim (n!)^2 1536\pi^{-5/2} n^{-2n} 16^n \mathrm{e}^{2n} n^{-17/2} \\ \sim 1536\pi^{-3/2} n^{-15/2} 16^n \quad \text{as } n \to \infty . \end{aligned} \tag{10.37}$$

☒

Other methods can be applied to Gessel's generating function to obtain asymptotics of $u_k(n)$ for wider ranges of k (Odlyzko et al. 1995).

The above example obtains a good estimate because the remainder term in (10.30) is smaller than the main term by a factor of $|z|^{-1}$. Had it been smaller only by a factor of $|z|^{-1/2}$, the resulting estimate would have been worthless, and it would have been necessary to obtain a fuller asymptotic expansion of $U_4(z)$ or else use smoothness properties of the remainder term. This is due to the phenomenon, mentioned before, that crude absolute-value estimates in either Cauchy's theorem, or the elementary approaches of section 8, usually lose a factor of $n^{1/2}$ when estimating the nth coefficient.

The subtraction-of-singularities principle can be applied even when the generating functions seem to be more complicated than those of Example 10.9. If we consider the problem of that example, but with $k = 5$, then we find that

$$U_5(z) = 3\exp(10z)(5 \cdot 2^{13} \cdot \pi^{5/2} z^{25/2})^{-1}(1 + \mathrm{O}(|z|^{-1})) \tag{10.38}$$

as $|z| \to \infty$, with $|\mathrm{Arg}(z)| \leqslant 3\pi/8$, $U_5(-z) = U_5(z)$, and $U_5(z)$ is entire. We now need an entire function with known coefficients that grows as $\exp(10z)z^{-25/2}$. This is not difficult to obtain, as

$$I_0(10z)z^{-12} - \sum_{j=1}^{12} c_j z^{-j} \tag{10.39}$$

for suitable coefficients c_j has the desired properties.

10.3. The residue theorem and sums as integrals

Sometimes sums that are not easily handled by other methods can be converted to integrals that can be evaluated explicitly or estimated by the residue theorem. If $t(z)$ is a meromorphic function that has first-order poles at $z = a$, $a + 1, \ldots, b$, with $a \in \mathbb{Z}$, each with residue 1, then

$$\sum_{n=a}^{b} f(n) = \frac{1}{2\pi \mathrm{i}} \int_{\Gamma} f(z)t(z)\,\mathrm{d}z\ , \tag{10.40}$$

where Γ is a simple closed contour enclosing $a, a + 1, \ldots, b$, provided $f(z)$ is analytic inside Γ and $t(z)$ has no singularities inside Γ aside from the first-order poles at $a, a + 1, \ldots, b$. If $t(z)$ is chosen to have residue $(-1)^n$ at $z = n$, then we obtain

$$\sum_{n=a}^{b} (-1)^n f(n) = \frac{1}{2\pi \mathrm{i}} \int_{\Gamma} f(z)t(z)\,\mathrm{d}z\ . \tag{10.41}$$

A useful example is given by the formula

$$\sum_{k=0}^{n} \binom{n}{k} (-1)^k f(k) = \frac{(-1)^n n!}{2\pi \mathrm{i}} \int_{\Gamma} \frac{f(z)\,\mathrm{d}z}{z(z-1)\cdots(z-n)}\ . \tag{10.42}$$

The advantage of (10.40) and (10.41) is that the integrals can often be manipulated to give good estimates. This is especially valuable for alternating sums such as (10.41). An analytic function $f(z)$ is extremely regular, so a sum such as that in (10.40) can often be estimated by methods such as the Euler–Maclaurin summation formula (section 5.3). However, that formula cannot always be applied to alternating sums such as that of (10.41), because the sign change destroys the regularity of $f(n)$. (However, as is noted in section 5.3, there are generalizations of the Euler–Maclaurin formula that are sometimes useful.) It is hard to write down general rules for applying this method, as most situations require appropriate choice of $t(z)$ and careful handling of the integral. For a detailed discussion of this method, often referred to as Rice's method, see section 4.9 of Henrici (1974–86). A pair of popular functions to use as $t(z)$ are

$$t_1(z) = \pi/(\sin \pi z), \qquad t_2(z) = \pi/(\tan \pi z) . \tag{10.43}$$

One can show (Theorem 4.9a of Henrici 1974–86) that if $r(z) = p(z)/q(z)$ with $p(z)$ and $q(z)$ polynomials such that $\deg q(z) \geqslant \deg p(z) + 2$, and $q(n) \neq 0$ for any $n \in \mathbb{Z}$, then

$$\sum_{n=-\infty}^{\infty} r(n) = -\sum \operatorname{Res}(r(z)t_1(z)) , \tag{10.44}$$

$$\sum_{n=-\infty}^{\infty} (-1)^n r(n) = -\sum \operatorname{Res}(r(z)t_2(z)) , \tag{10.45}$$

where the sums on the right-hand sides above are over the zeros of $q(z)$.

Examples of applications of these methods to asymptotics of data structures are given in Flajolet and Sedgewick (1986) and Szpankowski (1988).

10.4. Location of singularities, Rouché's theorem, and unimodality

A recurrent but only implicit theme throughout the discussion in this section is that of isolation of zeros. For example, to apply Theorem 9.2 we need to know that the polynomial $h(z)$ has only k zeros, each of multiplicity one, in $|z| < R$. Proofs of such results can often be obtained with the help of Rouché's theorem (Henrici 1974–86, Titchmarsh 1939).

Theorem 10.10. *Suppose that $f_1(z)$ and $f_2(z)$ are functions that are analytic inside and on the boundary of a simple closed contour Γ. If*

$$|f_2(z)| < |f_1(z)| \quad \textit{for all } z \in \Gamma , \tag{10.46}$$

then $f_1(z)$ and $f_1(z) + f_2(z)$ have the same number of zeros (counted with multiplicity) inside Γ.

Example 10.11 (*Sequences with forbidden subblocks*). We consider again the topic of Examples 6.4, 6.8, and 9.3, and prove the results that have already been used in Example 9.3. We again set

$$h(z) = z^k + (1 - 2z)C_A(z) , \tag{10.47}$$

where the only fact about $C_A(z)$ we will use is that it is a polynomial of degree $< k$ and coefficients 0 and 1, and $C_A(0) = 1$. We wish to show that $h(z)$ has only one zero in $|z| \leqslant 0.6$ if k is large. Write

$$C_A(z) = 1 + \frac{1}{2}\sum_{j=1}^{\infty} z^j + \frac{1}{2}\sum_{j=1}^{\infty} \varepsilon_j z^j \,, \tag{10.48}$$

where $\varepsilon_j = \pm 1$ for each j. Then

$$C_A(z) = \frac{2-z}{2(1-z)} + u(z) \,, \tag{10.49}$$

where

$$|u(z)| \leqslant \frac{|z|}{2(1-|z|)} \,.$$

For $|z| = r < 1$, we have $|u(z)| \leqslant r/(2(1-r))$. On the other hand, $z \to (2-z)/(1-z)$ maps circles to circles, since it is a fractional linear transformation, so it takes the circle $|z| = r$ to the circle with center on the real axis that goes through the two points $(2-r)/(1-r)$ and $(2+r)/(1+r)$. Therefore for $|z| = r < 1$,

$$|C_A(z)| \geqslant \frac{2+r}{2(1+r)} - \frac{r}{2(1-r)} = \frac{1-r-r^2}{1-r^2} \,, \tag{10.50}$$

and so $|C_A(z)| \geqslant 1/16$ for $|z| = r \leqslant 0.6$. Hence, if $k \geqslant 9$, then on $|z| = 0.6$,

$$|(1-2z)C_A(z)| \geqslant 1/80 > (0.6)^k \,, \tag{10.51}$$

and thus $(1-2z)C_A(z)$ and $h(z)$ have the same number of zeros in $|z| \leqslant 0.6$. On the other hand, $C_A(z)$ has no zeros in $|z| \leqslant 0.6$ by (10.50), while $1-2z$ has one, so we obtain the desired result, at least for $k \geqslant 9$. (A more careful analysis shows that $h(z)$ has only one root inside $|z| = 0.6$ even for $4 \leqslant k < 9$. For $1 \leqslant k \leqslant 3$, there are cases where there is no zero inside $|z| \leqslant 0.6$.) Example 6.7 shows how to obtain precise estimates of the single zero.

We note that (10.50) shows that for $|z| = 0.55$, $k \geqslant 9$

$$|h(z)| \geqslant |1 - 1.1|0.2 - (0.55)^k \geqslant 0.02 - 0.01 \geqslant 1/100 \,, \tag{10.52}$$

a result that was used in Example 9.3. ⊠

Example 10.12 (*Coins in a fountain*). An (n,k) fountain is an arrangement of n coins in rows such that there are k coins in the bottom row, and such that each coin in a higher row touches exactly two coins in the next lower row. Let $a_{n,k}$ be the number of (n,k) fountains, and $a_n = \sum_k a_{n,k}$ the total number of fountains of n coins. The values of a_n for $1 \leqslant n \leqslant 6$ are $1, 1, 2, 3, 5, 9$. If we let $a_0 = 1$ then it can be shown (Odlyzko and Wilf 1988) that

$$f(z) = \sum_{n=0}^{\infty} a_n z^n = \cfrac{1}{1 - \cfrac{z}{1 - \cfrac{z^2}{1 - \cfrac{z^3}{1 \cdots}}}} \,. \tag{10.53}$$

This is a famous continued fraction of Ramanujan. [Other combinatorial interpretations of this continued fraction are also known, see the references in Odlyzko and Wilf (1988). For related results, see Privman and Svrakic (1988, 1989).) Although one can derive the asymptotics of the a_n from the expansion (10.53), it is more convenient to work with another expansion, known from previous studies of Ramanujan's continued fraction:

$$f(z) = \frac{p(z)}{q(z)} , \tag{10.54}$$

where

$$p(z) = \sum_{r \geqslant 0} (-1)^r \frac{z^{r(r+1)}}{(1-z)(1-z^2)\cdots(1-z^r)} , \tag{10.55}$$

$$q(z) = \sum_{r \geqslant 0} (-1)^r \frac{z^{r^2}}{(1-z)(1-z^2)\cdots(1-z^r)} . \tag{10.56}$$

Clearly both $p(z)$ and $q(z)$ are analytic in $|z| < 1$, so $f(z)$ is meromorphic there. We will show that $q(z)$ has a simple real zero x_0, $0.57 < x_0 < 0.58$, and no other zeros in $|z| < 0.62$, while $p(x_0) > 0$. It will then follow from Theorem 10.8 that

$$a_n = cx_0^{-n} + \mathrm{O}((5/3)^n) \quad \text{as } n \to \infty , \tag{10.57}$$

where $c = -p(x_0)/(x_0 q'(x_0))$. Numerical computation shows that $c = 0.312\,36\ldots$, $x_0 = 0.576\,148\,769\ldots$.

To establish the claim about x_0, let $p_n(z)$ and $q_n(z)$ denote the nth partial sums of the series (10.55) and (10.56), respectively. Write $a(z) = q_3(z)(1-z)(1-z^2)/(1-z^3)$, so that

$$a(z) = 1 - 2z - z^2 + z^3 + 3z^4 + z^5 - 2z^6 - z^7 - z^9 , \tag{10.58}$$

and consider

$$b(z) = \prod_{j=1}^{9} (z - z_j) ,$$

where the z_j are $0.575\,77$, $-0.469\,97 \pm i0.817\,92$, $0.748\,33 \pm i0.075\,23$, $-1.059\,26 \pm i0.367\,18$, $0.493\,01 \pm i1.581\,85$, in that order. (The z_j are approximations to the zeros of $a(z)$, obtained from numerical library subroutines. How they were derived is not important for the verification of our proof.) An easy hand or machine computation shows that if $a(z) = \sum_k a_k z^k$, $b(z) = \sum b_k z^k$, then

$$\sum_{k=0}^{9} |a_k - b_k| \leqslant 1.7 \times 10^{-4} ,$$

and so $|a(z) - b(z)| \leqslant 1.7 \times 10^{-4}$ for all $|z| \leqslant 1$. Another computation shows that $|b(z)| \geqslant 8 \times 10^{-4}$ for all $|z| = 0.62$.

On the other hand, for $0 \leqslant u \leqslant 0.62$ and $|z| = u$, we have for $k \geqslant 5$

$$\left|\frac{z^{(k+1)^2-k^2}}{1-z^{k+1}}\right| \leqslant \frac{u^{2k+1}}{1-u^{k+1}} \leqslant \frac{u^9}{1-u^5}\,. \tag{10.59}$$

Therefore

$$\left|\sum_{k=4}^{\infty}(-1)^k \frac{z^{k^2}}{\prod_{j=4}^{k}(1-z^j)}\right| \leqslant \frac{u^{16}}{1-u^4}\sum_{m\geqslant 0}\left(\frac{u^9}{1-u^5}\right)^m \leqslant 6\times 10^{-4}\,, \tag{10.60}$$

and so by Rouché's theorem, $q(z)$ and $b(z)$ have the same number of zeros in $|z| \leqslant 0.62$, namely 1. Since $q(z)$ has real coefficients, its zero is real. This establishes the existence of x_0. An easy computation shows that $q(0.57) > 0$, $q(0.58) < 0$, so $0.57 < x_0 < 0.58$.

To show that $p(x_0) > 0$, note that successive summands in (10.55) decrease in absolute magnitude for each fixed real $z > 0$, and $p(z) > 1 - z^2/(1-z) > 0$ for $0 < z < 0.6$. ☒

The method used in the above example is widely applicable to generating functions given by continued fractions. Typically they are meromorphic in a disk centered at the origin, with a single dominant pole of order 1. Usually there is no convenient representation of the form (10.54) with explicit $p(z)$ and $q(z)$, and one has to work harder to establish the necessary properties about location of poles.

It was mentioned in section 6.4 that unimodality of a sequence is often deduced from information about the zeros of the associated polynomial. If the zeros of the polynomial

$$A(z) = \sum_{k=0}^{n} a_k z^k$$

are real and $\leqslant 0$, then the a_k are unimodal, and even the $a_k\binom{n}{k}^{-1}$ are log-concave. However, weaker properties follow from weaker assumptions on the zeros. If all the zeros of $A(z)$ are in the wedge-shaped region centered on the negative real axis $|\mathrm{Arg}(-z)| \leqslant \pi/4$, and the a_k are real, then the a_k are log-concave, but the $a_k\binom{n}{k}^{-1}$ are not necessarily log-concave. (This follows by factoring $A(z)$ into polynomials with real coefficients that are either linear or quadratic, and noting that all have log-concave coefficients, so their product does too.) One can prove other results that allow zeros to lie in larger regions, but then it is necessary to impose restrictions on ratios of their distances from the origin.

10.5. Implicit functions

Section 6.2 presented functions, such as $f^{(-1)}(z)$, that are defined implicitly. In this section we consider related problems that arise when a generating function $f(z)$ satisfies a functional equation $f(z) = G(z, f(z))$. Such equations arise frequently in graphical enumeration, and there is a standard procedure invented by Pólya and developed by Otter that is almost algorithmic (Harary and Palmer 1973, Harary et al. 1975) and routinely leads to them. Typically $G(z,w)$ is analytic in z and w in a small neighborhood of $(0,0)$. Zeros of analytic functions in more than one dimension are not isolated, and by the implicit function theorem $G(z,w) = w$ is solvable for w as a function of z, except for those points where

$$G_w(z,w) = \frac{\partial}{\partial w} G(z,w) = 1 \ . \tag{10.61}$$

Usually for z in a small neighborhood of 0 the solution w of $G(z,w) = w$ will not satisfy (10.61), and so w will be analytic in that neighborhood. As we enlarge the neighborhood under consideration, though, a simultaneous solution to $G(z,w) = w$ and (10.61) will eventually appear, and will usually be the dominant singularity of $f(z) = w(z)$. The following theorem covers many common enumeration problems.

Theorem 10.13. *Suppose that*

$$f(z) = \sum_{n=1}^{\infty} f_n z^n \tag{10.62}$$

is analytic at $z = 0$, that $f_n \geqslant 0$ for all n, and that $f(z) = G(z, f(z))$, where

$$G(z,w) = \sum_{m,n \geqslant 0} g_{m,n} z^m w^n \ . \tag{10.63}$$

Suppose that there exist real numbers $\delta, r, s > 0$ such that

(i) *$G(z,w)$ is analytic in $|z| < r + \delta$ and $|w| < s + \delta$,*

(ii) *$G(r,s) = s$, $G_w(r,s) = 1$,*

(iii) *$G_z(r,s) \neq 0$ and $G_{ww}(r,s) \neq 0$.*

Suppose that $g_{m,n} \in \mathbb{R}^+ \cup \{0\}$ for all m and n, $g_{0,0} = 0$, $g_{0,1} = 1$, and $g_{m,n} > 0$ for some m and some $n \geqslant 2$. Assume further that there exist $h > j > i \geqslant 1$ such that $f_h f_i f_j \neq 0$ while the greatest common divisor of $j - i$ and $h - i$ is 1. Then $f(z)$ converges at $z = r$, $f(r) = s$, and

$$f_n = [z^n] f(z) \sim (r G_z(r,s)/(2\pi G_{ww}(r,s)))^{1/2} n^{-3/2} r^{-n} \quad \text{as } n \to \infty \ . \tag{10.64}$$

Example 10.14 (*Rooted labeled trees*). As was shown in Example 6.1, the exponential generating function $t(z)$ of rooted labeled trees satisfies $t(z) = z \exp(t(z))$. Thus we have $G(z,w) = z \exp(w)$, and Theorem 10.13 is easily seen to apply with $r = \mathrm{e}^{-1}$, $s = 1$. Therefore we obtain the asymptotic estimate

$$t_n/n! = [z^n] t(z) \sim (2\pi)^{-1/2} n^{-3/2} \mathrm{e}^n \quad \text{as } n \to \infty \ . \tag{10.65}$$

On the other hand, from Example 6.6 we know that $t_n = n^{n-1}$, a much more satisfactory answer, so that the estimate (10.65) only provides us with another proof of Stirling's formula. ⊠

The example above involves an extremely simple application of Theorem 10.13. More complicated cases will be presented in section 15.1.

The statement of Theorem 10.13 is long, and the hypotheses stringent. All that is really needed for the asymptotic relation (10.64) to hold is that $f(z)$ should be analytic on $\{z\colon |z| \leqslant r, z \neq r\}$ and that

$$f(z) = c(r - z)^{1/2} + o(|r - z|^{1/2}) \tag{10.66}$$

for $|z - r| \leqslant \varepsilon$, $|\mathrm{Arg}(r - z)| \geqslant \pi/2 - \varepsilon$ for some $\varepsilon > 0$. If these conditions are satisfied, then (10.64) follows immediately from either the transfer theorems of section 11.1 or (with stronger hypotheses) from Darboux's method of section 11.2. The purpose of Theorem 10.13 is to present a general theorem that guarantees (10.66) holds, is widely applicable, and is stated to the maximum extent possible in terms of conditions on the coefficients of $f(z)$ and $G(z, w)$.

Theorem 10.13 is based on Theorem 5 of Bender (1974) and Theorem 1 of Meir and Moon (1989). The hypotheses of Bender's Theorem 5 are simpler than those of Theorem 10.13, but, as was pointed out by Canfield (1984), the proof is faulty and there are counterexamples to the claims of that theorem. The difficulty is that Theorem 5 of Bender (1974) does not distinguish adequately between the different solutions $w = w(z)$ of $w = G(z, w)$, and the singularity of the combinatorially significant solution may not be the smallest among all singularities of all solutions. The result of Meir and Moon (1989) provides conditions that assure such pathological behavior does not occur. [The statement of Theorem 10.13 incorporates some corrections to Theorem 1 of Meir and Moon (1989) provided by the authors of that paper.] It would be desirable to prove results like (10.64) under a simpler set of conditions.

In many problems the function $G(z, w)$ is of the form

$$G(z, w) = g(z)\phi(w) + h(z)\ , \tag{10.67}$$

where $g(z)$, $\phi(w)$, and $h(z)$ are analytic at 0. For this case Meir and Moon (1989) have proved a useful result (their Theorem 2) that implies an asymptotic estimate of the type (10.64). The hypotheses of that result are often easier to verify than those of Theorem 10.13 above. [As was noted by Meir and Moon (1989), the last part of their conditions (4.12a) has to be replaced by the condition that $y_i > h_i$, $y_j > h_j$, and $y_k > h_k$ for some $k > j > i \geqslant 1$ with $\gcd(j - i, k - i) = 1$.]

Whenever Theorem 10.13 applies, $f_n = [z^n]f(z)$ equals the quantity on the right-hand side of (10.64) to within a multiplicative factor of $1 + O(n^{-1})$. One can derive fuller expansions for the ratio when needed.

11. Small singularities of analytic functions

In most combinatorial enumeration applications, the generating function has a single dominant singularity. The methods used to extract asymptotic information about coefficients split naturally into two main classes, depending on whether this singularity is large or small.

In some situations the same generating function can be said to have either a large or a small singularity, depending on the range of coefficients that we are interested in. This is illustrated by the following example.

Example 11.1 (*Partitions with bounded part sizes*). Let $p(n,m)$ be the number of (unordered) partitions of an integer n into integers $\leqslant m$. It is easy to see that

$$P_m(z) = \sum_{n=0}^{\infty} p(n,m)z^n = \prod_{k=1}^{m} (1-z^k)^{-1} . \tag{11.1}$$

The function $P_m(z)$ is rational, but has to be treated in different ways depending on the relationship of n and m. If n is large compared to m, it turns out to be appropriate to say that $P_m(z)$ has a small singularity, and use methods designed for this type of problems. However, if n is not too large compared to m, then the singularity of $P_m(z)$ can be said to be large. [Since the largest part in a partition of n is almost always $O(n^{1/2}\log n)$ (Erdős and Lehner 1941), $p(n,m) \sim p(n)$ if m is much larger than $n^{1/2}\log n$.]

Although $P_m(z)$ has singularities at all the kth roots of unity for all $k \leqslant m$, $z = 1$ is clearly the dominant singularity, as $|P_m(r)|$ grows much faster as $r \to 1^-$ than $|P_m(z)|$ for $z = r\exp(i\theta)$ for any $\theta \in (0, 2\pi)$. If m is fixed, then the partial function decomposition can be used to obtain the asymptotics of $p(n,m)$ as $m \to \infty$. We cannot use Theorem 9.2 directly, since the pole of $P_m(z)$ at $z = 1$ has multiplicity 1. However, either by using the generalizations of Theorem 9.2 that are mentioned in section 9.1, or by the subtraction-of-singularities principle, we can show that for any fixed m,

$$\begin{aligned} p(n,m) &\sim [z^n] \left(\prod_{k=1}^{m} k! \right)^{-1} (1-z)^{-m} \\ &\sim \left(\prod_{k=1}^{m} k! \right)^{-1} ((m-1)!)^{-1} \quad \text{as } n \to \infty . \end{aligned} \tag{11.2}$$

[See Ayoub (1963) for further details and estimates.] This approach can be extended for m growing slowly with n, and it can be shown without much effort that the estimate (11.2) holds for $n \to \infty$, $m \leqslant \log\log n$, say. However, for larger values of m this approach becomes cumbersome, and other methods, such as those of section 12, are necessary. ⊠

11.1. Transfer theorems

This section presents some results, drawn from Flajolet and Odlyzko (1990b), that allow one to translate an asymptotic expansion of a generating function around its dominant singularity into an asymptotic expansion for the coefficients in a direct way. These results are useful in combinatorial enumeration, since the conditions for validity are frequently satisfied. The proofs, which we do not present here, are based on the subtraction-of-singularities principle, but are more involved than the cases treated in section 10.2.

We start out with an application of the results to be presented later in this section.

Example 11.2 (*2-regular graphs*). The generating function for 2-regular graphs is known (Comtet 1974) to be

$$f(z) = (1-z)^{-1/2}\exp\left(-\frac{1}{2}z - \frac{1}{4}z^2\right) . \tag{11.3}$$

[A simpler proof can be obtained from the exponential formula, cf. eq. (3.9.1) of Wilf (1990).] We see that $f(z)$ is analytic throughout the complex plane except for the slit along the real axis from 1 to ∞, and that near $z = 1$ it has the asymptotic expansion

$$f(z) = \mathrm{e}^{-3/4}\left\{(1-z)^{-1/2} + (1-z)^{1/2} + \frac{1}{4}(1-z)^{3/2} + \cdots\right\} . \tag{11.4}$$

Theorem 11.4 below then shows that as $n \to \infty$,

$$\begin{aligned}[z^n]f(z) &\sim \mathrm{e}^{-3/4}\left\{\binom{n-1/2}{n} + \binom{n-3/2}{n} + \frac{1}{4}\binom{n-5/2}{n} + \cdots\right\} \\ &\sim \frac{\mathrm{e}^{-3/4}}{\sqrt{\pi n}}\left\{1 - \frac{5}{8n} - \frac{15}{128n^2} + \cdots\right\} .\end{aligned} \tag{11.5}$$

☒

The basic transfer results will be presented for generating functions that have a single dominant singularity, but can be extended substantially beyond their circle of convergence. For r, $\eta > 0$, and $0 < \phi < \pi/2$, we define the closed domain $\Delta = \Delta(r, \phi, \eta)$ by

$$\Delta(r, \phi, \eta) = \{z\colon |z| \leqslant r + \eta,\ |\mathrm{Arg}(z - r)| \geqslant \phi\} . \tag{11.6}$$

In the main result below we will assume that a generating function is analytic throughout $\Delta \setminus \{r\}$. Later in this section we will mention some results that dispense with this requirement. We will also explain why analyticity throughout $\Delta \setminus \{r\}$ is helpful in obtaining results such as those of Theorem 11.4 below.

One advantage to using Cauchy's theorem to recover information about coefficients of generating functions is that it allows one to prove the intuitively obvious result that small smooth changes in the generating function correspond to small smooth changes in the coefficients. We will use the quantitative notion of a function of slow variation at ∞ to describe those functions for which this notion can be made precise. (With more effort one can prove that the same results hold with a less restrictive definition than that below.)

Definition 11.3. A function $L(u)$ is of *slow variation* at ∞ if

(i) There exist real numbers u_0 and ϕ_0 with $u_0 > 0$, $0 < \phi_0 < \pi/2$, such that $L(u)$ is analytic and $\neq 0$ in the domain

$$\{u\colon |\operatorname{Arg}(u-u_0)| \leqslant \pi - \phi_0\} . \tag{11.7}$$

(ii) There exists a function $\varepsilon(x)$, defined for $x \geqslant 0$ with $\lim_{x\to\infty} \varepsilon(x) = 0$, such that for all $\theta \in [-(\pi - \phi_0), \pi - \phi_0]$ and $u \geqslant u_0$, we have

$$\left| \frac{L(u\,\mathrm{e}^{\mathrm{i}\theta})}{L(u)} - 1 \right| < \varepsilon(u) \tag{11.8}$$

and

$$\left| \frac{L(u \log^2 u)}{L(u)} - 1 \right| < \varepsilon(u) . \tag{11.9}$$

Theorem 11.4. *Assume that $f(z)$ is analytic throughout the domain $\Delta \setminus \{r\}$, where $\Delta = \Delta(r, \phi, \eta)$, $r, \eta > 0$, $0 < \phi < \pi/2$, and that $L(u)$ is a function of slow variation at ∞. If α is any real number, then*

(A) *If*

$$f(z) = \mathrm{O}\left((z-r)^\alpha L\left(\frac{1}{r-z}\right)\right)$$

uniformly for $z \in \Delta \setminus \{r\}$, then

$$[z^n]f(z) = \mathrm{O}(r^{-n} n^{-\alpha-1} L(n)) \quad \text{as } n \to \infty .$$

(B) *If*

$$f(z) = \mathrm{o}\left((z-r)^\alpha L\left(\frac{1}{r-z}\right)\right)$$

uniformly as $z \to r$ for $z \in \Delta \setminus \{r\}$, then

$$[z^n]f(z) = \mathrm{o}(r^{-n} n^{-\alpha-1} L(n)) \quad \text{as } n \to \infty .$$

(C) *If $\alpha \notin \{0, 1, 2, \ldots\}$ and*

$$f(z) \sim (r-z)^\alpha L\left(\frac{1}{r-z}\right)$$

uniformly as $z \to r$ for $z \in \Delta \setminus \{r\}$, then

$$[z^n]f(z) \sim \frac{r^{-n} n^{-\alpha-1}}{\Gamma(-\alpha)} L(n) .$$

The restriction that there be only one singularity on the circle of convergence is easy to relax. If there are several (corresponding to oscillatory behavior of the coefficients), their contributions to the coefficients add. The crucial fact is that at each singularity the function $f(z)$ should be continuous except for an angular region similar to that of $\Delta(r, \phi, \eta)$.

The requirement that the generating function $f(z)$ be analytic in the interior of $\Delta(r, \phi, \eta)$ is in general harder to dispense with, at least by the methods of Flajolet and Odlyzko (1990b). However, if the singularity at r is sufficiently large, one can obtain the same results with weaker assumptions that only require analyticity inside the disk $|z| < r$. The following result is implicit in Flajolet and Odlyzko (1990b).

Theorem 11.5. *Assume that $f(z)$ is analytic in the domain$\{z\colon |z| \leqslant r, z \neq r\}$ and that $L(u)$ is a function of slow variation at ∞. If α is any fixed real number with $\alpha < -1$, then the implications* (A), (B), *and* (C) *of Theorem* 11.4 *are valid.*

Example 11.6 (*Longest cycle in a random permutation*). The average length of the longest cycle in a permutation on n letters is $[z^n]f(z)$, where

$$f(z) = (1-z)^{-1} \sum_{k \geqslant 0} \left[1 - \exp\left(- \sum_{j \geqslant k} j^{-1} z^j \right) \right] .$$

It is easy to see that $f(z)$ is analytic in $|z| < 1$, and a double application of the Euler–Maclaurin summation formula shows that $f(z) \sim G(1-z)^{-2}$ as $z \to 1$, uniformly for $|z| \leqslant 1$, $z \neq 1$, where

$$G = \int_0^\infty \left[1 - \exp\left(- \int_x^\infty t^{-1} \mathrm{e}^{-t} \, \mathrm{d}t \right) \right] \mathrm{d}x = 0.624 \ldots . \tag{11.10}$$

Therefore, by Theorem 11.5 with $L(u) = 1$,

$$[z^n]f(z) \sim Gn \quad \text{as } n \to \infty , \tag{11.11}$$

a result first proved by Shepp and Lloyd (1966) using Poisson approximations and Tauberian theorems. The derivation sketched above follows Flajolet and Odlyzko (1990a,b). Flajolet and Odlyzko (1990a) contains many other applications of transfer theorems to random mapping problems. Additional recent papers on the cycle structure of random permutations are Arratia and Tavaré (1992a) and Hansen (1994). They use probabilistic methods, not transfer theorems, and contain extensive references to other recent works. ☒

In applying transfer theorems, it is useful to have explicit expansions and estimates for the coefficients of some frequently occurring functions. We state several asymptotic series:

$$[z^n](1-z)^\alpha \approx \frac{n^{-\alpha-1}}{\Gamma(-\alpha)} \left(1 + \sum_{k \geqslant 1} e_k^{(\alpha)} n^{-k} \right), \quad \alpha \neq 0, 1, 2, \ldots, \tag{11.12}$$

where

$$e_k^{(\alpha)} = \sum_{j=k}^{2k} (-1)^j \lambda_{k,j} (\alpha+1)(\alpha+2) \cdots (\alpha+j) , \tag{11.13}$$

and the $\lambda_{k,j}$ are determined by

$$\mathrm{e}^t (1 + vt)^{-1-1/v} = \sum_{k,j \geqslant 0} \lambda_{k,j} v^k t^j . \tag{11.14}$$

In particular,

$$e_1^{(\alpha)} = \alpha(\alpha+1)/2,$$
$$e_2^{(\alpha)} = \alpha(\alpha+1)(\alpha+2)(3\alpha+1)/24\,.$$

Also, for $\alpha, \beta \notin \{0,1,2,\dots\}$,

$$[z^n](1-z)^\alpha(-z^{-1}\log(1-z))^\beta \approx \frac{n^{-\alpha-1}}{\Gamma(-\alpha)}(\log n)^\beta\left(1+\sum_{k\geqslant 1} e_k^{(\alpha,\beta)}(\log n)^{-k}\right), \tag{11.15}$$

where

$$e_k^{(\alpha,\beta)} = (-1)^k\binom{\beta}{k}\Gamma(-\alpha)\left(\frac{\mathrm{d}^k}{\mathrm{d}s^k}\Gamma(-s)^{-1}\bigg|_{s=\alpha}\right)\,. \tag{11.16}$$

Further examples of asymptotic expansions are presented in Flajolet and Odlyzko (1990b).

Why is the analyticity of a function $f(z)$ throughout $\Delta(r,\phi,\eta)\setminus\{r\}$ so important? We explain this using as an example a function $f(z)$ that satisfies

$$f(z) = (1+\mathrm{o}(1))(1-z)^{1/2} \tag{11.17}$$

as $z \to 1$ with $z \in \Delta = \Delta(1, \pi/8, 1)$. We write

$$f(z) = (1-z)^{1/2} + g(z)\,, \tag{11.18}$$

so that

$$|g(z)| = \mathrm{o}(|1-z|^{1/2})\,. \tag{11.19}$$

Since $[z^n](1-z)^{1/2}$ grows like $n^{-3/2}$, we would like to show that

$$|[z^n]g(z)| = \mathrm{o}(n^{-3/2}) \quad \text{as } n \to \infty\,. \tag{11.20}$$

If $g(z)$ were analytic in a disk of radius $1+\delta$ for some $\delta > 0$, then we could conclude that $|[z^n]g(z)| < (1+\delta/2)^{-n}$ for large n, a conclusion much stronger than (11.20). However, if all we know is that $g(z)$ satisfies (11.19) in $|z| \leqslant 1$, then we can only conclude from Cauchy's theorem that $[z^n]g(z) = \mathrm{O}(1)$, since (11.19) implies that $|g(z)| \leqslant C$ for all $|z| < 1$ and some $C > 0$. Then Theorem 10.2 gives

$$|[z^n]g(z)| \leqslant Cr^{-n} \tag{11.21}$$

uniformly for all $n \geqslant 0$ and all $r < 1$, and hence $|[z^n]g(z)| \leqslant C$ for all n, a result that is far from what is required. If we know that $g(z)$ can be continued to $\Delta\setminus\{r\}$ and satisfies (11.19) there, we can do a lot better. We choose the contour $\Gamma = \Gamma_1 \cup \Gamma_2 \cup \Gamma_3 \cup \Gamma_4$, pictured in fig. 1, with

$$\Gamma_1 = \{z\colon |z-1| = 1/n,\ |\mathrm{Arg}(z-1)| \geqslant \pi/4\}\,, \tag{11.22}$$

$$\Gamma_2 = \{z\colon z = 1 + r\exp(\pi\mathrm{i}/4),\ 1/n \leqslant r \leqslant \delta\}\,, \tag{11.23}$$

$$\Gamma_3 = \{z\colon |z| = |1+\delta\exp(\pi\mathrm{i}/4)|,\ |\mathrm{Arg}(z-1)| \geqslant \pi/4\}\,, \tag{11.24}$$

$$\Gamma_4 = \{z\colon z = 1 + r\exp(-\pi\mathrm{i}/4),\ 1/n \leqslant r \leqslant \delta\}\,, \tag{11.25}$$

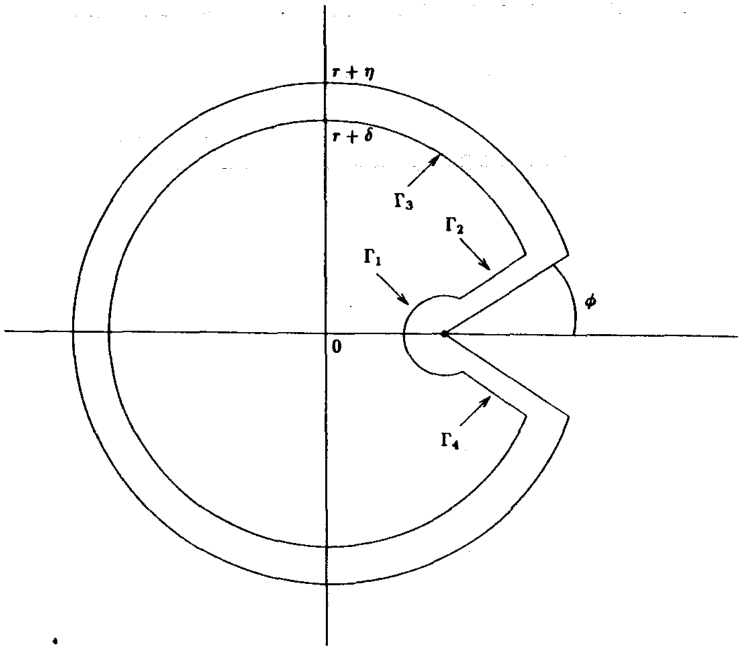


Figure 1. Domain $\Delta(r, \phi, \eta)$ of section 11.1 and the integration contour Γ.

where $0 < \delta < 1/2$. We will show that the integrals

$$g_j = \frac{1}{2\pi i} \int_{\Gamma_j} g(z) z^{-n-1} \, dz \tag{11.26}$$

on the Γ_j are small. On Γ_3, $g(z)$ is bounded, so we trivially obtain the exponential upper bound

$$|g_3| = O((1 + \delta/2)^{-n}) \, . \tag{11.27}$$

On Γ_1, $|g(z)| = o(n^{-1/2})$, $|z^{-n-1}| \leqslant (1 - 1/n)^{-n-1} = O(1)$, and the length of Γ_1 is $\leqslant 2\pi/n$, so

$$|g_1| = o(n^{-3/2}) \quad \text{as } n \to \infty \, . \tag{11.28}$$

Next, on Γ_2, for $z = 1 + r\exp(\pi i/4)$,

$$\begin{aligned} |z|^{-n} &= |1 + r2^{-1/2} + ir2^{-1/2}|^{-n} = (1 + r2^{1/2} + r^2)^{-n/2} \\ &\leqslant (1 + r)^{-n/2} \leqslant \exp(-nr/10) \end{aligned} \tag{11.29}$$

for $0 \leqslant r < 1$. Since $g(z)$ satisfies (11.19), for any $\varepsilon > 0$ we have

$$|g(1 + r\exp(\pi i/4))| \leqslant \varepsilon r^{1/2} \tag{11.30}$$

if $0 < r \leqslant \eta$ for some $\eta = \eta(\varepsilon) \leqslant \delta$. Therefore

$$|g_2| \leqslant \varepsilon \int_0^{\eta} r^{1/2} \exp(-nr/10)\,\mathrm{d}r + \mathrm{O}\left(\int_{\eta}^{\infty} \exp(-nr/10)\,\mathrm{d}r\right)$$
$$\leqslant \varepsilon n^{-3/2} \int_0^{\infty} r^{1/2} \exp(-r/10)\,\mathrm{d}r + \mathrm{O}(\exp(-n\eta/10))\ , \tag{11.31}$$

and so

$$|g_2| = \mathrm{o}(n^{-3/2})\ . \tag{11.32}$$

Since $|g_4| = |g_2|$, inequalities (11.27), (11.28), and (11.32) show that (11.20) holds.

The critical factor in the derivation of (11.20) was the bound (11.29) for $|z|^{-n}$ on the segment $z = 1 + r\exp(\pi\mathrm{i}/4)$. Integrating on the circle $|z| = 1$ or even on the line $\mathrm{Re}\,(z) = 1$ does not give a bound for $|z|^{-n}$ that is anywhere as small, and the resulting bounds do not approach (11.20) in strength. The use of the circular arc Γ_1 in the integral is only a minor technical device used to avoid the singularity at $z = 1$.

When one cannot continue a function to a region like $\Delta \setminus \{1\}$, it is sometimes possible to obtain good estimates for coefficients by working with the generating function exclusively in $|z| \leqslant 1$, provided some smoothness properties apply. This method is outlined in the next section.

11.2. Darboux's theorem and other methods

A singularity of $f(z)$ at $z = w$ is called algebraic if $f(z)$ can be written as the sum of a function analytic in a neighborhood of w and a finite number of terms of the form

$$(1 - z/w)^{\alpha} g(z)\ , \tag{11.33}$$

where $g(z)$ is analytic near w, $g(w) \neq 0$, and $\alpha \notin \{0, 1, 2, \ldots\}$. Darboux's theorem (Darboux 1878) gives asymptotic expansions for functions with algebraic singularities on the circle of convergence. We state one form of Darboux's result, derived from Theorem 8.4 of Szegö (1959).

Theorem 11.7. *Suppose that $f(z)$ is analytic for $|z| < r$, $r > 0$, and has only algebraic singularities on $|z| = r$. Let a be the minimum of* $\mathrm{Re}\,(\alpha)$ *for the terms of the form* (11.33) *at the singularities of $f(z)$ on $|z| = r$, and let w_j, α_j, and $g_j(z)$ be the w, α, and $g(z)$ for those terms of the form* (11.33) *for which* $\mathrm{Re}\,(\alpha) = a$. *Then, as $n \to \infty$,*

$$[z^n]f(z) - \sum_j \frac{g_j(w_j) n^{-\alpha_j - 1}}{\Gamma(-\alpha_j) w_j^n} + \mathrm{o}(r^{-n} n^{-a-1})\ . \tag{11.34}$$

Jungen (1931) has extended Darboux's theorem to functions that have a single dominant singularity which is of a mixed algebraic and logarithmic form. His method can be applied also to functions that have several such singularities on their circle of convergence.

We do not devote much attention to Darboux's and Jungen's theorems because they can be obtained from the transfer theorems of section 11.1. The only reason for stating Theorem 11.7 is that it occurs frequently in the literature.

Some functions, such as

$$f(z) = \prod_{k=1}^{\infty}(1 + z^k/k^2) , \tag{11.35}$$

are analytic in $|z| \leqslant 1$, cannot be continued outside the unit circle, yet are nicely behaved on $|z| = 1$. Therefore there is no dominant singularity that can be studied to determine the asymptotics of $[z^n]f(z)$. To minimize the size of the integrand, it is natural to move the contour of integration in Cauchy's formula to the unit circle. Once that is done, it is possible to exploit smoothness properties of $f(z)$ to bound the coefficients. The Riemann–Lebesgue lemma implies that if $f(z)$ is integrable on the unit circle, then as $n \to \infty$,

$$[z^n]f(z) = (2\pi)^{-1} \int_{-\pi}^{\pi} f(\mathrm{e}^{\mathrm{i}\theta}) \exp(-n\mathrm{i}\theta)\, \mathrm{d}\theta = \mathrm{o}(1) . \tag{11.36}$$

More can be said if the derivative of $f(z)$ exists on the unit circle. When we apply integration by parts to the integral in (11.36), we find

$$[z^n]f(z) = (2\pi n)^{-1} \int_{-\pi}^{\pi} f'(\mathrm{e}^{\mathrm{i}\theta}) \exp(-(n-1)\mathrm{i}\theta)\, \mathrm{d}\theta , \tag{11.37}$$

and so $|[z^n]f(z)| = \mathrm{o}(n^{-1})$ if $f'(z)$ exists and is integrable on the unit circle. Existence of higher derivatives leads to even better estimates. We do not attempt to state a general theorem, but illustrate an application of this method with an example. The same technique can be used in other situations, for example in obtaining better error terms in Darboux's theorem (Darboux 1878).

Example 11.8 (*Permutations with distinct cycle lengths*). Example 8.10 showed that for the function $f(z)$ defined by eq. (8.58), $[z^n]f(z) \sim \exp(-\gamma)$ as $n \to \infty$. This coefficient is the probability that a random permutation on n letters has distinct cycle lengths. The more precise estimate (8.59) was derived by Greene and Knuth (1982) by working with recurrences for the coefficients of $f(z)$ and auxiliary functions. Another approach to deriving fuller asymptotic expansions for $[z^n]f(z)$ is to use the method outlined above. It suffices to show that the function $g(z)$ defined by eq. (8.62) has a nice expansion in the closed disk $|z| \leqslant 1$. Since

$$g(z) = -z + \sum_{m=2}^{\infty} \frac{(-1)^{m-1}}{m} \{\mathrm{Li}_m(z^m) - z^m\} , \tag{11.38}$$

where the $\mathrm{Li}_m(w)$ are the polylogarithm functions (Lewin 1981), one can use the theory of the $\mathrm{Li}_m(w)$. A simpler way to proceed is to note, for example, that

$$\sum_{k=2}^{\infty} \frac{z^{2k}}{k^2} = \sum_{k=2}^{\infty} \frac{z^{2k}}{k(k-1)} + r(z) , \tag{11.39}$$

where

$$r(z) = -\sum_{k=2}^{\infty} \frac{z^{2k}}{k^2(k-1)}, \tag{11.40}$$

and so $r'(z)$ is bounded and continuous for $|z| \leqslant 1$, as are the terms in (8.62) with $m \geqslant 3$. On the other hand,

$$\sum_{k=2}^{\infty} \frac{z^{2k}}{k(k-1)} = z^2 + (1-z^2)\log(1-z^2); \tag{11.41}$$

so we can write $g(z) = g_1(z) + g_2(z)$, where $g_1(z)$ is an explicit function [given by eq. (11.41)] such that the coefficients of $\exp(g_1(z))$ can be estimated asymptotically using transfer methods or other techniques, and $g_2(z)$ has the property that $g_2'(z)$ is bounded and continuous in $|z| \leqslant 1$. Continuing this process, we can find, for every K, an expansion for the coefficients of $f(z)$ that has error term $O(n^{-K})$. To do this, we write $g(z) = G_1(z) + G_2(z)$. In this expansion $G_1(z)$ will be explicitly given and analytic inside $|z| < 1$ and analytically continuable to some region that extends beyond the unit disk with the exception of cuts from a finite number of points on the unit circle out to infinity. Further, $G_2(z)$ will have the property that $G_2^{(K)}(z)$ is bounded and continuous in $|z| \leqslant 1$. This will then give the desired expansion for the coefficients of $f(z)$. ☒

12. Large singularities of analytic functions

This section presents methods for asymptotic estimation of coefficients of generating functions whose dominant singularities are large.

12.1. The saddle point method

The saddle point method, also referred to as the method of steepest descent, is by far the most useful method for obtaining asymptotic information about rapidly growing functions. It is extremely flexible and has been applied to a tremendous variety of problems. It is also complicated, and there is no simple categorization of situations where it can be applied, much less of the results it produces. Given the purpose and limitations on the length of this chapter, we do not present a full discussion of it. For a complete and insightful introduction to this technique, the reader is referred to de Bruijn (1958). Many other books, such as Evgrafov (1961), Fedoryuk (1989a), Olver (1974), and Wong (1989) also have extensive presentations. What this section does is to outline the method and show when and how it can be applied and what kinds of estimates it produces. Examples of proper and improper applications of the method are presented. Later subsections are then devoted to general results obtained through applications of the saddle point method. These results give asymptotic expansions for wide classes of functions without forcing the reader to go through the details of the saddle point method.

The saddle point method is based on the freedom to shift contours of integration when estimating integrals of analytic functions. The same principle underlies other techniques, such as the transfer method of section 11.1, but the way it is applied here is different. When dealing with functions of slow growth near their principal singularity, as happens for transfer methods, one attempts to push the contour of integration up to and in some ways even beyond the singularity. The saddle point method is usually applied when the singularity is large, and it keeps the path of integration close to the singularity.

In the remainder of this section we will assume that $f(z)$ is analytic in $|z| < R \leqslant \infty$. We will also make the assumption that for some R_0, if $R_0 < r < R$, then

$$\max_{|z|=r} |f(z)| = f(r) . \tag{12.1}$$

This assumption is clearly satisfied by all functions with real nonnegative coefficients, which are the most common ones in combinatorial enumeration. Further, we will suppose that $z = r$ is the unique point with $|z| = r$ where the maximum value in (12.1) is assumed. When this assumption is not satisfied, we are almost always dealing with some periodicity in the asymptotics of the coefficients, and we can then usually reduce to the standard case by either changing variables or rewriting the generating function as a sum of several others, as was discussed in section 10. [Such a reduction cannot be applied to the function of eq. (9.39), though.]

The first step in estimating $[z^n]f(z)$ by the saddle point method is to find the saddle point. Under our assumptions, that will be a point $r \in (R_0, R)$ which minimizes $r^{-n}f(r)$. We have encountered this condition before, in section 8.1. The minimizing $r = r_0$ will usually be unique, at least for large n. (If there are several $r \in (R_0, R)$ for which $r^{-n}f(r)$ achieves its minimum value, then $f(z)$ is pathological, and the standard saddle point method will not be applicable. For functions $f(z)$ with nonnegative coefficients, it is easy to show uniqueness of the minimizing r, as has already been discussed in section 8.1.) Cauchy's formula (10.6) is then applied with the contour $|z| = r_0$. The reason for this choice is that for many functions, on this contour the integrand is large only near $z = r_0$, the contributions from the region near $z = r_0$ do not cancel each other, and remaining regions contribute little. This is in contrast to the behavior of the integrand on other contours. By Cauchy's theorem, any simple closed contour enclosing the origin gives the correct answer. However, on most of them the integrand is large, and there is so much cancellation that it is hard to derive any estimates. The circle going through the saddle point, on the other hand, yields an integral that can be controlled well by techniques related to Laplace's method and the method of stationary phase that were mentioned in section 5.5. We illustrate with an example, which is a totally self-contained application of the saddle point method to an extremely simple situation.

Example 12.1 (*Stirling's formula*). We estimate $(n!)^{-1} = [z^n]\exp(z)$. The saddle point, according to our definition above, is that $r \in \mathbb{R}^+$ that minimizes $r^{-n}\exp(r)$, which is clearly $r = n$. Consider the contour $|z| = n$, and set $z = n\exp(\mathrm{i}\theta)$, $-\pi \leqslant$

$\theta \leqslant \pi$. Then

$$
\begin{aligned}
[z^n]\exp(z) &= \frac{1}{2\pi i}\int_{|z|=n} \frac{\exp(z)}{z^{n+1}}\,dz \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} n^{-n}\exp(n e^{i\theta} - ni\theta)\,d\theta\ . \qquad (12.2)
\end{aligned}
$$

Since $|\exp(z)| = \exp(\mathrm{Re}(z))$, the absolute value of the integrand in (12.2) is $n^{-n}\exp(n\cos\theta)$, which is maximized for $\theta = 0$. Now

$$
e^{i\theta} = \cos\theta + i\sin\theta = 1 - \theta^2/2 + i\theta + O(|\theta|^3)\ ,
$$

so for any $\theta_0 \in (0, \pi)$,

$$
\int_{-\theta_0}^{\theta_0} n^{-n}\exp(n e^{i\theta} - ni\theta)\,d\theta = \int_{-\theta_0}^{\theta_0} n^{-n}\exp(n - n\theta^2/2 + O(n|\theta|^3))\,d\theta\ . \qquad (12.3)
$$

(It is the cancellation of the $ni\theta$ term coming from $n e^{i\theta}$ and the $-ni\theta$ term that came from change of variables in z^{-n} that is primarily responsible for the success of the saddle point method.) The $O(n|\theta|^3)$ term in (12.3) could cause problems if it became too large, so we will select $\theta_0 = n^{-2/5}$, so that $n|\theta|^3 \leqslant n^{-1/5}$ for $|\theta| \leqslant \theta_0$, and therefore

$$
\exp(n - n\theta^2/2 + O(n|\theta|^3)) = \exp(n - n\theta^2/2)(1 + O(n^{-1/5}))\ . \qquad (12.4)
$$

Hence

$$
\int_{-\theta_0}^{\theta_0} n^{-n}\exp(n e^{i\theta} - ni\theta)\,d\theta = (1 + O(n^{-1/5}))n^{-n}e^n \int_{-\theta_0}^{\theta} \exp(-n\theta^2/2)\,d\theta\ .
$$

But

$$
\begin{aligned}
\int_{-\theta_0}^{\theta_0} \exp(-n\theta^2/2)\,d\theta &= \int_{-\infty}^{\infty} \exp(-n\theta^2/2)\,d\theta - 2\int_{\theta_0}^{\infty} \exp(-n\theta^2/2)\,d\theta \\
&= (2\pi/n)^{1/2} - O(\exp(-n^{1/5}/2))\ ,
\end{aligned}
$$

so

$$
\int_{-\theta_0}^{\theta_0} n^{-n}\exp(n e^{i\theta} - ni\theta)\,d\theta = (1 + O(n^{-1/5}))(2\pi/n)^{1/2}n^{-n}e^n\ . \qquad (12.5)
$$

On the other hand, for $\theta_0 < |\theta| \leqslant \pi$,

$$
\cos\theta \leqslant \cos\theta_0 = 1 - \theta_0^2/2 + O(\theta_0^4)\ ,
$$

so

$$
n\cos\theta \leqslant n - n^{1/5}/2 + O(n^{-3/5})\ ,
$$

and therefore for large n

$$\left| \int_{\theta_0}^{\pi} n^{-n} \exp(n\,\mathrm{e}^{\mathrm{i}\theta} - n\mathrm{i}\theta)\,\mathrm{d}\theta \right| \leqslant n^{-n} \exp(n - n^{1/5}/3) ,$$

and similarly for the integral from $-\pi$ to $-\theta_0$. Combining all these estimates we therefore find that

$$(n!)^{-1} = [z^n] \exp(z) = (1 + \mathrm{O}(n^{-1/5}))(2\pi n)^{-1/2} n^{-n}\, \mathrm{e}^n , \tag{12.6}$$

which is a weak form of Stirling's formula (4.3). (The full formula can be derived by using more precise expansions for the integrand.)

Suppose we try to push through a similar argument using the contour $|z| = 2n$. This time, instead of eq. (12.2), we find

$$[z^n] \exp(z) = \frac{1}{2\pi} \int_{\pi}^{\pi} 2^{-n} n^{-n} \exp(2n\,\mathrm{e}^{\mathrm{i}\theta} - n\mathrm{i}\theta)\,\mathrm{d}\theta . \tag{12.7}$$

At $\theta = 0$, the integrand is $2^{-n}n^{-n}\exp(2n)$, which is $\exp(n)$ times as large as the value of the integrand in (12.2). Since the two integrals do produce the same answer, and from the analysis above we see that this answer is close to $n^{-n}\exp(n)$ in value, the integral in (12.7) must involve tremendous cancellation. That is indeed what we see in the neighborhood of $\theta = 0$. We find that

$$\exp(2n\,\mathrm{e}^{\mathrm{i}\theta} - n\mathrm{i}\theta) = \exp(2n - n\theta^2 + n\mathrm{i}\theta + \mathrm{O}(n|\theta|^3)) , \tag{12.8}$$

and the $\exp(n\mathrm{i}\theta)$ term produces wild oscillations of the integrand even over small ranges of θ. Trying to work with the integral (12.7) and proving that it equals something exponentially smaller than the maximal value of its integrand is not a promising approach. By contrast, the saddle point contour used to produce eq. (12.2) gives nice behavior of the integrand, so that it can be evaluated. ☒

The estimates for $n!$ obtained in Example 10.1 came from a simple application of the saddle point method. The motivation for the choice of the contour $|z| = n$ is provided by the discussion at the end of the example; other choices lead to oscillating integrands that cannot be approximated by a Gaussian, nor by any other nice function. The example above treated only the exponential function, but it is easy to see that this phenomenon is general; a rapidly oscillating term $\exp(n\mathrm{i}\alpha)$ for $\alpha \neq 0$ is present unless the contour passes through the saddle point. When we do use this contour, and the Gaussian approximation is valid, we find that for functions $f(z)$ satisfying our assumptions we have the following estimate.

Saddle point approximation.

$$[z^n]f(z) \sim (2\pi b(r_0))^{-1/2} f(r_0) r_0^{-n} \quad \text{as } n \to \infty , \tag{12.9}$$

where r_0 is the saddle point (where $r^{-n}f(r)$ is minimized, so that $r_0 f'(r_0)/f(r_0) = n$) and

$$b(r) = r\frac{f'(r)}{f(r)} + r^2\frac{f''(r)}{f(r)} - r^2\left(\frac{f'(r)}{f(r)}\right)^2 = r\left(r\frac{f'(r)}{f(r)}\right)' . \tag{12.10}$$

Example 12.2 (*Bell numbers*). Example 5.4 showed how to estimate the Bell number B_n by elementary methods, starting with the representation (5.38). The exponential generating function

$$B(z) = \sum_{n=0}^{\infty} B_n \frac{z^n}{n!} \tag{12.11}$$

satisfies

$$B(z) = \exp(\exp(z) - 1) ,$$

as can be seen from (5.38) or by other methods (cf. Comtet 1974). The saddle point occurs at that $r_0 > 0$ that satisfies

$$r_0 \exp(r_0) = n , \tag{12.12}$$

and

$$b(r_0) = r_0(1 + r_0) \exp(r_0) , \tag{12.13}$$

so the saddle point approximation says that as $n \to \infty$,

$$B_n \sim n!(2\pi r_0^2 \exp(r_0))^{-1/2} \exp(\exp(r_0) - 1) r_0^{-n} . \tag{12.14}$$

The saddle point approximation can be justified even more easily than for the Stirling estimate of $n!$. ☒

The above approximation is widely applicable and extremely useful, but care has to be exercised is applying it. This is shown by the next example.

Example 12.3 (*Invalid application of the saddle point method*). Consider the trivial example $f(z) = (1 - z)^{-1}$, so that $[z^n]f(z) = 1$ for all $n \geqslant 0$. Then $f'(r)/f(r) = (1 - r)^{-1}$, and so the saddle point is $r_0 = n/(n + 1)$, and $b(r_0) = r_0/(1 - r_0)^2 = n(n + 1)$. Therefore if the approximation (12.9) were valid, it would give

$$\begin{aligned} [z^n]f(z) &\sim (2\pi n(n + 1))^{-1/2}(n + 1)\left(1 + \frac{1}{n}\right)^n \\ &\sim (2\pi)^{-1/2} \mathrm{e} \quad \text{as } n \to \infty . \end{aligned} \tag{12.15}$$

Since $(2\pi)^{-1/2}\mathrm{e} = 1.0844\ldots \neq 1 = [z^n]f(z)$, something is wrong, and the estimate (12.9) does not apply to this function. ☒

The estimate (12.9) gave the wrong result in Example 12.3 because the Gaussian approximation on the saddle-point-method contour used so effectively in Example 12.1 (and in almost all cases where the saddle point method applies) does not hold over a sufficiently large region for $f(z) = (1 - z)^{-1}$. In Example 12.1 we used without detailed explanation the choice $\theta_0 = n^{-2/5}$, which gave the approximation (12.5) for $|\theta| \leqslant \theta_0$, and yet led to an estimate for the integral over $\theta_0 < |\theta| \leqslant \pi$ that was negligible. This was possible because the third-order term (i.e., $n|\theta|^3$) in eq. (12.5) was small. When we try to imitate this approach for $f(z) = (1-z)^{-1}$, we fail, because the third-order term is too large. Instead of $n\,\mathrm{e}^{\mathrm{i}\theta} - n\mathrm{i}\theta$, we now have

$$-\log(1-r_0\,\mathrm{e}^{\mathrm{i}\theta}) - n\mathrm{i}\theta = -\log(1-r_0) - \frac{1}{2}n(n+1)\theta^2 - \frac{\mathrm{i}}{6}n^2(n+1)\theta^3 + \cdots . \tag{12.16}$$

More fundamentally, the saddle point method fails here because the function $f(z) = (1-z)^{-1}$ does not have a large enough singularity at $z = 1$, so that when one traverses the saddle point contour $|z| = r_0$, the integrand does not drop off rapidly enough for a small region near the real axis to provide the dominant contribution.

When can one apply the saddle point approximation (12.9)? Perhaps the simplest, yet still general, set of sufficient conditions for the validity of (12.9) is provided by requiring that the function $f(z)$ be Hayman-admissible. Hayman admissibility is described in Definition 12.4, in the following subsection. Generally speaking, though, for the saddle point method to apply we need the function $f(z)$ to have a large dominant singularity at R, so that $f(r)$ grows at least as fast as $\exp((\log(R-r))^2)$ as $r \to R^-$ for $R < \infty$, and as fast as $\exp((\log r)^2)$ as $r \to \infty$ for $R = \infty$. The faster the growth rate, the easier it usually is to apply the method, so that $\exp(1/(1-z))$ or $\exp(\exp(1/(1-z)))$ can be treated easily.

In our application of the saddle point method to $\exp(z)$ in Example 12.1 we were content to obtain a poor error term, $1 + \mathrm{O}(n^{-1/5})$, in Stirling's formula for $n!$. This was done to simplify the presentation and concentrate only on the main factors that make the saddle point method successful. With more care devoted to the integral one can obtain the full asymptotic expansion of $n!$. (Only the range $|\theta| \leqslant \theta_0$ has to be considered carefully.) This is usually true when the saddle point method is applicable.

This section provided a sketchy introduction to the saddle point method. For a much more thorough presentation, including a discussion of the topographical view of the integrand and the "hill-climbing" interpretation of the contour of integration, see de Bruijn (1958).

12.2. Admissible functions

The saddle point method is a powerful and flexible tool, but in its full generality it is often cumbersome to apply. In many situations it is possible to apply general theorems derived using the saddle point method that give asymptotic approximations that are not the sharpest possible, but which allow one to avoid the drudgery of applying the method step by step. The general theorems that we present were proved by Hayman (1956) and by Harris and Schoenfeld (1968). We next describe the classes of functions to which these theorems apply, and then present the estimates one obtains for them. It is not always easy to verify that these definitions hold, but it is almost always easier to do this than to apply the saddle point method from scratch. It is worth mentioning, furthermore, that for many generating functions, there are conditions that guarantee that they satisfy the hypotheses of the Hayman and the Harris–Schoenfeld theorems. These conditions are discussed later in this section.

The definition below is stated somewhat differently than the original one in Hayman (1956), but can be shown to be equivalent to it.

Definition 12.4. A function

$$f(z)=\sum_{n=0}^{\infty} f_n z^n \tag{12.17}$$

is admissible in the sense of Hayman (or *H-admissible*) if

(i) $f(z)$ is analytic in $|z|<R$ for some $0<R\leqslant\infty$,

(ii) $f(z)$ is real for z real, $|z|<R$,

(iii) for $R_0<r<R$,

$$\max_{|z|=r}|f(z)|=f(r)\,, \tag{12.18}$$

(iv) for

$$a(r)=r\,\frac{f'(r)}{f(r)}\,, \tag{12.19}$$

$$b(r)=ra'(r)\;=\;r\,\frac{f'(r)}{f(r)}+r^2\frac{f''(r)}{f(r)}-r^2\left(\frac{f'(r)}{f(r)}\right)^2\,, \tag{12.20}$$

and for some function $\delta(r)$, defined in the range $R_0<r<R$ to satisfy $0<\delta(r)<\pi$, the following three conditions hold:

(a) $f(r\,\mathrm{e}^{\mathrm{i}\theta})\sim f(r)\,\exp(\mathrm{i}\theta a(r)-\theta^2 b(r)/2)$

as $r\to R$ uniformly for $|\theta|<\delta(r)$, (12.21)

(b) $f(r\,\mathrm{e}^{\mathrm{i}\theta})=\mathrm{o}(f(r)b(r)^{-1/2})$

as $r\to R$ uniformly for $|\theta|<\delta(r)$, (12.22)

(c) $b(r)\to\infty$ as $r\to R$. (12.23)

For H-admissible functions, Hayman (1956) proved a basic result that gives the asymptotics of the coefficients.

Theorem 12.5. *If $f(z)$, defined by eq.* (12.17), *is H-admissible in $|z|<R$, then*

$$f_n=(2\pi b(r))^{-1/2}f(r)r^{-n}\left\{\exp\left(-\frac{(a(r)-n)^2}{b(r)}\right)+\mathrm{o}(1)\right\} \tag{12.24}$$

as $r\to R$, with the o(1) *term uniform in n.*

If we choose $r=r_n$ to be a solution to $a(r_n)=n$, then we obtain from Theorem 12.5 a simpler result. (The uniqueness of r_n follows from a result of Hayman (1956) which shows that $a(r)$ is positive increasing in some range $R_1<r<R$, $R_1>R_0$.)

Corollary 12.6. *If $f(z)$, defined by eq. (12.17), is H-admissible in $|z| < R$, then*

$$f_n \sim (2\pi b(r_n))^{-1/2} f(r_n) r_n^{-n} \quad \text{as } n \to \infty \,, \tag{12.25}$$

where r_n is defined uniquely for large n by $a(r_n) = n$, $R_0 < r_n < R$.

Corollary 12.6 is adequate for most situations. The advantage of Theorem 12.5 is that it gives a uniform estimate over the approximate range $|a(r) - n| \lesssim b(r)^{1/2}$. [Note that the estimate (12.24) is vacuous for $|a(r) - n| \; b(r)^{-1/2} \to \infty$.] Theorem 12.5 shows that the $f_n r^n$ are approximately Gaussian in the central region.

There are many direct applications of the above results.

Example 12.7 (*Stirling's formula*). Let $f(z) = \exp(z)$. Then $f(z)$ is H-admissible for $R = \infty$; conditions (i)–(iii) of Definition 12.4 are trivially satisfied, while $a(r) = r$, $b(r) = r$, so (iv) also holds for $R_0 = 0$, $\delta(r) = r^{-1/3}$, say. Corollary 12.6 then shows that

$$f_n = \frac{1}{n!} \sim (2\pi n)^{-1/2} \, \mathrm{e}^n n^{-n} \quad \text{as } n \to \infty \,, \tag{12.26}$$

since $r_n = n$, which gives a weak form of Stirling's approximation to $n!$. ☒

In many situations the conditions of H-admissibility are much harder to verify than for $f(z) = \exp(z)$, and even in that case there is a little work to be done to verify that condition (iv) holds. However, many of the generating functions one encounters are built up from other, simpler generating functions, and Hayman (1956) has shown that often the resulting functions are guaranteed to be H-admissible. We summarize some of Hayman's results in the following theorem.

Theorem 12.8. *Let $f(z)$ and $g(z)$ be H-admissible for $|z| < R \leqslant \infty$. Let $h(z)$ be analytic in $|z| < R$ and real for real z. Let $p(z)$ be a polynomial with real coefficients.*

(i) *If the coefficients a_n of the Taylor series of* $\exp(p(z))$ *are positive for all sufficiently large n, then* $\exp(p(z))$ *is H-admissible in $|z| < \infty$.*

(ii) $\exp(f(z))$ *and $f(z)g(z)$ are H-admissible in $|z| < R$.*

(iii) *If, for some $\eta > 0$, and $R_1 < r < R$,*

$$\max_{|z|=r} |h(z)| = O(f(r)^{1-\eta}) \,, \tag{12.27}$$

then $f(z) + h(z)$ is H-admissible in $|z| < R$. In particular, $f(z) + p(z)$ is H-admissible in $|z| < R$ and, if the leading coefficient of $p(z)$ is positive, $p(f(z))$ is H-admissible in $|z| < R$.

Example 12.9 (*H-admissible functions*). (a) By (i) Theorem 12.8, $\exp(z)$ is H-admissible, so we immediately obtain the estimate (12.26), which yields Stirling's formula. (b) Since $\exp(z)$ is H-admissible, part (iii) of Theorem 12.8 shows that $\exp(z) - 1$ is H-admissible. (c) Applying part (ii) of Theorem 12.8, we next find that $\exp(\exp(z) - 1)$ is H-admissible, which yields the asymptotics of the Bell numbers. ☒

Hayman's results give only first-order approximations for the coefficients of H-admissible functions. In some circumstances it is desirable to obtain full asymptotic expansions. This is possible if we impose additional restrictions on the generating function. We next state some results of Harris and Schoenfeld (1968).

Definition 12.10. A function $f(z)$ defined by eq. (12.17) is *HS-admissible* provided it is analytic in $|z| < R$, $0 < R \leqslant \infty$, is real for real x, and satisfies the following conditions:

(A) There is an R_0, $0 < R_0 < R$ and a function $d(r)$ defined for $r \in (R_0, R)$ such that

$$0 < d(r) < 1 \ , \qquad r\{1 + d(r)\} < R \ , \tag{12.28}$$

and such that $f(z) \neq 0$ for $|z - r| < rd(r)$.

(B) If we define, for $k \geqslant 1$,

$$A(z) = \frac{f'(z)}{f(z)}, \quad B_k(z) = \frac{z^k}{k!} A^{(k-1)}(z), \quad B(z) = \frac{z}{2} B_1(z) \ , \tag{12.29}$$

then we have

$$B(r) > 0 \ \text{ for } R_0 < r < R \ \text{ and } \ B_1(r) \to \infty \ \text{ as } r \to R \ .$$

(C) For sufficiently large R_1 and n, there is a unique solution $r = u_n$ to

$$B_1(r) = n + 1, \quad R_1 < r < R \ . \tag{12.30}$$

Let

$$C_j(z,r) = \frac{-1}{B(r)} \left\{ B_{j+2}(z) + \frac{(-1)^j}{j+2} B_1(r) \right\} \ . \tag{12.31}$$

There exist nonnegative D_n, E_n, and n_0 such that for $n \geqslant n_0$,

$$|C_j(u_n, u_n)| \leqslant E_n D_n^j, \quad j = 1, 2, \ldots \ . \tag{12.32}$$

(D) As $n \to \infty$,

$$\begin{aligned} &B(u_n) d(u_n)^2 \to \infty \ , \\ &D_n E_n B(u_n) d(u_n)^3 \to 0 \ , \\ &D_n d(u_n) \to 0 \ . \end{aligned} \tag{12.33}$$

For HS-admissible functions, Harris and Schoenfeld obtain complete asymptotic expansions.

Theorem 12.11. *If $f(z)$, defined by (12.17), is HS-admissible, then for any $N \geqslant 0$,*

$$f_n = 2(\pi\beta_n)^{-1/2} f(u_n) u_n^{-n} \left\{ 1 + \sum_{k=1}^{N} F_k(n) \beta_n^{-k} + \mathrm{O}(\phi_N(n; d)) \right\} \ \textit{as } n \to \infty \ , \tag{12.34}$$

where

$$\beta_n = B(u_n)\ , \tag{12.35}$$

$$F_k(n) = \frac{(-1)^k}{\sqrt{\pi}} \sum_{m=1}^{2k} \frac{\Gamma(m+k+\frac{1}{2})}{m!} \sum_{\substack{j_1+\cdots+j_m=2k \\ j_1,\ldots,j_m \geqslant 1}} \gamma_{j_1}(n)\cdots\gamma_{j_m}(n)\ , \tag{12.36}$$

$$\gamma_j(n) = C_j(u_n, u_n)\ , \tag{12.37}$$

and

$$\phi_N(n;d) = \max\{\mu(u_n,d), E_n'(D_nE_n''\beta_n^{-1/2})^{2N+2}\}\ ,$$

with

$$E_n' = \min(1,E_n),\quad E_n'' = \max(1,E_n)\ , \tag{12.38}$$

$$\mu(r,d) = \max\left\{\lambda(r;d)B(r)^{1/2},\ \frac{\exp(-B(r)d(r)^2)}{d(r)B(r)^{1/2}}\right\}\ , \tag{12.39}$$

where $\lambda(r;d)$ is the maximum value of $|f'(z)/f(z)|$ for z on the oriented path $Q(r)$ consisting of the line segment from $r+\mathrm{i}rd(r)$ to $(1-d(r)^2)^{1/2}+\mathrm{i}rd(r)$ and of the circular arc from the last point to $\mathrm{i}r$ to $-r$.

The conditions for HS-admissibility are often hard to verify. However, there is a theorem (Odlyzko and Richmond 1985c) which guarantees that they do hold for a large class of interesting functions.

Theorem 12.12. *If $g(z)$ is H-admissible, then $f(z) = \exp(g(z))$ is HS-admissible. Furthermore, the error term $\phi_N(n;d)$ of Theorem* 12.11 *is then* $\mathrm{o}(\beta_n^{-N})$ *as $n \to \infty$ for every fixed $N \geqslant 0$.*

Example 12.13 (*Bell numbers and HS-admissibility*). Since $\exp(x) - 1$ is H-admissible, as we saw in Example 12.9, we find that $\exp(\exp(z) - 1)$ is HS-admissible, and Theorem 12.11 yields a complete asymptotic expansion of the Bell numbers. ⊠

Theorem 12.12 does not apply when $g(z)$ is a polynomial. As is pointed out by Schmutz (1989), for $g(z) = z^4 - z^3 + z^2$ the function $f(z) = \exp(g(z))$ is HS-admissible, but Theorem 12.11 does not give an asymptotic expansion because the error term $\phi_N(n;d)$ is too large. Schmutz (1989) has obtained necessary and sufficient conditions for Theorem 12.11 to give an asymptotic expansion for the coefficients of $f(z) = \exp(g(z))$ when $g(z)$ is a polynomial.

12.3. Other saddle point applications

Section 12.1 presented the basic saddle point method and discussed its range of applicability. Section 12.2 was devoted to results derived using this method that are general and yet can be applied in a cookbook style, without a deep understanding of the saddle point technique. Such a cookbook approach is satisfactory in many situations. However, often one encounters asymptotic estimation problems that are not covered by any of general results mentioned in section 12.2, but can be solved using the saddle point method. This section mentions several such results of this type that illustrate the range of problems to which this method is applicable. Additional applications will be presented in section 15, where other techniques are combined with the saddle point method.

Example 12.14 (*Stirling numbers*). The Stirling numbers of the first kind, $s(n,k)$, satisfy (6.5) as well as (Comtet 1974):

$$\sum_{k=0}^{n} s(n,k)z^k = z(z-1)\cdots(z-n+1)\,. \tag{12.40}$$

Since $(-1)^{n+k}s(n,k) > 0$, [which is reflected in the behavior of the generating function (12.40), which grows faster along the negative real axis than along the positive one], we rewrite it as

$$\sum_{k=0}^{n} (-1)^{n+k}s(n,k)z^k = z(z+1)\cdots(z+n-1)\,. \tag{12.41}$$

The function on the right-hand side behaves like a good candidate for an application of the saddle point method. For details, see Moser and Wyman (1958a,b). ☒

The estimates mentioned in Example 12.14 are far from best possible in either the size of the error term or (more important) in the range of validity. References for the best currently known results about Stirling numbers of both the first and second kind are given in Temme (1993). Some of the results in the literature are not rigorous. For example, Temme (1993) presents elegant and uniform estimates based on an application of the saddle point method. They are likely to be correct, but the necessary rigorous error analysis has not been performed yet, although it seems that this should be doable. Other results, like those of Knessl and Keller (1991) are obtained by methods that there does not seem to be any hope of making rigorous in the near future. Some of the results, though, such as the original ones of Moser and Wyman (1958a,b), and the more recent one of Wilf (1993), are fully proved.

The saddle point method can be used to obtain full asymptotic expansions. These expansions are usually in powers of $n^{-1/2}$ when estimating $[z^n]f(z)$, and they hardly ever converge, but are asymptotic expansions as defined by Poincaré [as in eq. (2.2)]. The usual forms of the saddle point method are incapable of providing expansions similar to the Hardy–Ramanujan–Rademacher convergent series for the partition function $p(n)$ [eq. (3.1)]. However, the saddle point method can be applied to estimate $p(n)$. There are technical difficulties, since the generating function

$$f(z) = \sum_{n=0}^{\infty} p(n)z^n = \prod_{k=1}^{\infty}(1-z^k)^{-1} \tag{12.42}$$

has a large singularity at $z = 1$, but in addition has singularities at all other roots of unity. The contribution of the integral for z away from 1 can be crudely estimated to be $O(n^{-1}\exp(Cn^{1/2}/2))$ [the last term in eq. (1.5)]. A simple estimate of the integral near $z = 1$ yields the asymptotic expansion of eq. (1.6). A more careful treatment of the integral, but one that follows the conventional saddle point technique, replaces the $1 + O(n^{-1/2})$ term in eq. (1.6) by an asymptotic (in the sense of Poincaré, so nonconvergent) series $\sum c_k n^{-k/2}$. To obtain eq. (1.5), one needs to choose the contour of integration near $z = 1$ carefully and use precise estimates of $f(z)$ near $z = 1$.

De Bruijn (1958) also discusses applications of the saddle point method when the saddle point is not on the real axis, and especially when there are several saddle points that contribute comparable amounts. This usually occurs when there are oscillations in the coefficients. When the oscillations are irregular, the tricks mentioned in section 10 of changing variables do not work, and the contributions of the multiple saddle points have to be evaluated.

Example 12.15 (*Oscillating sequence*). Consider the sequence a_n of Examples 9.5 and 10.3. As is shown in Example 9.5, its ordinary generating function is given by (9.39). It has an essential singularity at $z = 1$, but is analytic everywhere else. This function is not covered by our earlier discussion. For example, its maximal value is in general not taken on the positive real axis. It can be shown that the Cauchy integral has two saddle points, at approximately $z = 1 - (2n)^{-1} \pm in^{-1/2}(1 - (4n)^{-1})^{1/2}$. Evaluating $[z^n]f(z)$ by using Cauchy's theorem with the contour chosen to pass through the two points in the correct way yields the estimate (9.38). ☒

In applying the saddle point method, a general principle is that multiplying a generating function $f(z)$ with dominant singularity at R by another function $g(z)$ which is analytic in $|z| < R$ and has much lower growth rate near $z = R$ yields a function $f(z)g(z)$ whose saddle point is close to that of $f(z)$. Usually one can obtain a relation of the form

$$[z^n](f(z)g(z)) \sim g(r_0)([z^n]f(z)) , \tag{12.43}$$

where r_0 is the saddle point for $f(z)$. This principle (which is related to the one behind Theorem 7.1) is useful, but has to be applied with caution, and proofs have to be provided for each case. For fuller exposition of this principle and general results, see Gardy (1995). The advantage of this approach is that often $f(z)$ is easy to manipulate, so the determination of a saddle point for it is easy, whereas multiplying it by $g(z)$ produces a messy function, and the exact saddle point for $f(z)g(z)$ is difficult to determine.

Example 12.16 (*Boolean lattice of subsets of* $\{1,\ldots,n\}$). The number a_n of Boolean sublattices of the Boolean lattice of subsets of $\{1,\ldots,n\}$ has the exponential generating function (Getu et al. 1992):

$$A(z) = \sum_{n=0}^{\infty} a_n \frac{z^n}{n!} = \exp(2z + \exp(z) - 1) . \tag{12.44}$$

We can write $A(z) = \exp(2z)B(z)$, where $B(z)$ is the exponential generating function for the Bell numbers (ple 12.2). Since $B(z)$ grows much faster than $\exp(2z)$, it is easy to show that (12.43) applies, and so

$$a_n \sim \exp(2r_0)B_n \quad \text{as } n \to \infty , \tag{12.45}$$

where r_0 is the saddle point for $B(z)$. Using the approximation (12.12) of Example 12.2, we find that

$$a_n \sim (n/\log n)^2 B_n \quad \text{as } n \to \infty . \tag{12.46}$$

☒

The insensitivity of the saddle point approximation to slight perturbations is reflected in slightly different definitions of a saddle point that are used. The saddle point approximation (12.9) for $[z^n]f(z)$ is stated in terms of r_0, the point that minimizes $f(r)r^{-n}$. The discussion of the saddle point emphasized minimization of the peak value of the integrand in Cauchy's formula, which is the same as minimizing $f(r)r^{-n-1}$, since the contour integral (10.6) involves $f(z)z^{-n-1}$. Some sources call the point minimizing $f(r)r^{-n-1}$ the saddle point. It is not important which definition is adopted. The asymptotic-series coefficients look slightly differently in the two cases, but the final asymptotic series, when expressed in terms of n, are the same. The reason for slightly preferring the definition that minimizes $f(r)r^{-n}$ is that when the change of variable $z = r\exp(\mathrm{i}\theta)$ is made in Cauchy's integral, there is no linear term in θ, and the integrand involves $\exp(-cn\theta^2 + O(|\theta|^3))$. If we minimized $f(r)r^{-n-1}$, we would have to deal with $\exp(-c'\mathrm{i}\theta - c''n\theta^2 + O(|\theta|^3))$, which is not much more difficult to handle but is less elegant.

The same principle can be applied when the exact saddle point is hard to determine, and it is awkward to work with an implicit definition of this point. When that happens, there is often a point near the saddle point that is easy to handle, and for which the saddle point approximation holds. We refer to Gardy (1995) for examples and discussion of this phenomenon.

12.4. The circle method and other techniques

As we mentioned in section 12.3, the saddle point method is a powerful method that estimates the contribution of the neighborhood of only a single point, or at most a few points. The convergent series of eq. (1.3) for the partition function $p(n)$ (as well as the earlier nonconvergent but asymptotic and very accurate expansion of Hardy and Ramanujan) is obtained by evaluating the contribution of the other singularities of $f(z)$ to the integral. The mth term in eq. (1.3) comes from the primitive mth roots of unity. To obtain this expansion one needs to use a special contour of integration and detailed knowledge of the behavior of $f(z)$. The details of this technique, called the circle method, can be found in Andrews (1976) and Ayoub (1963).

Convergent series can be obtained from the circle method only when the generating function is of a special form. For results and references, see Almkvist (1991) and Almkvist and Andrews (1991).

Nonconvergent but accurate asymptotic expansions can be derived from the circle method in a much wider variety of applications. It is especially useful when there is no single dominant singularity. For the partition function $p(n)$, all the singularities away from $z = 1$ contribute little, and it is $z = 1$ that creates the dominant term and yields eq. (1.6). For other functions this is often false. For example, when dealing with additive problems of Waring's type, where one studies $N_{k,m}(n)$, the number of representations of a nonnegative integer n as

$$n = \sum_{j=1}^{m} x_j^k , \quad x_j \in \mathbb{Z}^+ \cup \{0\} \quad \text{for all } j , \tag{12.47}$$

the natural generating function to study is

$$\sum_{n=0}^{\infty} N_{k,m}(n) z^n = g(z)^m , \tag{12.48}$$

where

$$g(z) = \sum_{h=0}^{\infty} z^{h^k} . \tag{12.49}$$

The function $g(z)$ has a natural boundary at $|z| = 1$, but it again grows fastest as z approaches a root of unity from within $|z| < 1$, so it is natural to speak of $g(z)$ having singularities at the roots of unity. The singularity at $z = 1$ is still the largest, but not by much, as other roots of unity contribute comparable amounts, with the contribution of other roots of unity ζ diminishing as the order of ζ increases. All the contributions can be estimated, and one can obtain solutions to Waring's problem (which was to show that for every k, there is an integer m such that $N_{k,m}(n) > 0$ for all n) and other additive problems. For details of this method see Ayoub (1963). We mention here that for technical reasons, one normally works with generating functions of the form $G_n(z)^m$, where

$$G_n(z) = \sum_{h=0}^{\lfloor n^{1/k} \rfloor} z^{h^k} , \tag{12.50}$$

(so that the generating function depends on n), and analyzes them for $|z| = 1$ (since they are now polynomials), but the basic explanation above of why this process works still applies.

13. Multivariate generating functions

A major difficulty in estimating the coefficients of multivariate generating functions is that the geometry of the problem is far more difficult. It is harder to see what are the critical regions where the behavior of the function determines the asymptotics of the coefficients, and those regions are more complicated. Singularities and zeros are no longer isolated, as in the one-dimensional case, but instead form $(k-1)$-dimensional manifolds in k variables. Even rational multivariate functions are not easy to deal with.

One basic tool in one-dimensional complex analysis is the residue theorem, which allows one to move a contour of integration past a pole of the integrand. (We derived a form of the residue theorem in section 10, in the discussion of poles of generating functions.) There is an impressive generalization by Leray (Aizenberg and Yuzhakov 1983, Leray 1959) of this theory to several dimensions. Unfortunately, it is complicated, and with few exceptions [such as that of Lichtin (1991), see also Bertozzi and McKenna (1993)] so far it has not been applied successfully to enumeration problems. On the other hand, there are some much simpler tools that can frequently be used to good effect.

An important tool in asymptotics of multivariate generating functions is the multidimensional saddle point method.

Example 13.1 (*Alternating sums of powers of binomial coefficients*). Consider

$$S(s,n)=\sum_{k=0}^{2n}(-1)^{k+n}\binom{2n}{k}^{s}\,, \tag{13.1}$$

where s and n are positive integers. It has been known for a long time that $S(1,n)=0$, $S(2,n)=(2n)!(n!)^{-2}$, $S(3,n)=(3n)!(n!)^{-3}$. However, no formula of this type has been known for $s>3$. De Bruijn (1958, chapter 4) showed that $S(s,n)$ for integer $s>3$ cannot be expressed as a ratio of products of factorials. Although his proof is not presented as an application of the multidimensional saddle point method, it is easy to translate it into those terms. $S(s,n)$ is easily seen to equal the constant term in

$$F(z_1,\ldots,z_{s-1})=(-1)^n(1+z_1)^{2n}\cdots(1+z_{s-1})^{2n}(1-(z_1\cdots z_{s-1})^{-1})^{2n}\,, \tag{13.2}$$

and so

$$S(s,n)=(2\pi i)^{-s+1}\int\cdots\int F(z_1,\ldots,z_{s-1})z_1^{-1}\cdots z_{s-1}^{-1}\,dz_1\cdots dz_{s-1}\,, \tag{13.3}$$

where the integral is taken with each z_j traversing a circle, say. De Bruijn's proof in effect shows that for s fixed and $n \to \infty$, there are two saddle points at $z_1 = \cdots = z_{s-1} = \exp(2i\alpha)$, with $\alpha = \pm(2s)^{-1}$, and this leads to the estimate

$$S(s,n) \sim \left\{2\cos\left(\frac{\pi}{2s}\right)\right\}^{2ns+s-1} 2^{2-s}(\pi n)^{(1-s)/2}s^{-1/2} \text{ as } n \to \infty\,, \tag{13.4}$$

valid for any fixed integer $s \geqslant 2$. Since $\cos(\pi(2s)^{-1})$ is algebraic but irrational for $s \geqslant 4$, the asymptotic estimate (13.4) shows that $S(s,n)$ cannot be expressed as a ratio of finite products of $(a_j n)!$ for any fixed finite set of integers a_j.

De Bruijn (1958, chapter 6) derives the asymptotics of $S(s,n)$ as $n \to \infty$ for general real s. The approach sketched above no longer applies, and de Bruijn uses the integral representation

$$S(s,n) = \int_C \left(\frac{\Gamma(2n+1)}{\Gamma(n+z+1)\Gamma(n-z+1)}\right)^s \frac{\mathrm{d}z}{2\mathrm{i}\sin\pi z}\,,$$

where C is a simple closed curve that contains the points $-n, -n+1, \ldots, -1, 0, 1, \ldots, n$ in its interior and has no other integer points on the real axis in its closure. A complicated combination of analytic techniques, including the one-dimensional saddle point method, then leads to the final asymptotic estimate of $S(s,n)$. ☒

The multidimensional saddle point method works best when applied to large singularities. Just as for the basic one-dimensional method, it does not work when applied to small singularities, such as those of rational functions. Fortunately, there is a trick that often succeeds in converting a small singularity in n dimensions into a large one in $n-1$ dimensions. The main idea is to expand the generating function with respect to one of the variables through partial-fraction expansions or other methods. It is hard to write down a general theorem, but the next example illustrates this technique.

Example 13.2 (*Alignments of k sequences*). Let $f(k,n)$ denote the number of $k \times m$ matrices of 0's and 1's such that each column sum is $\geqslant 1$ and each row sum is exactly n. (The number of columns, m, can vary, although obviously $k \leqslant m \leqslant kn$.) We consider k fixed, $n \to \infty$ (Griggs et al. 1990). If we let $N(r_1, \ldots, r_k)$ denote the number of 0, 1 matrices with k rows, no columns of all 0's, and row sums $r_1, \ldots, r_k$, then it is easy to see (Griggs et al. 1990) that

$$F(z_1, \ldots, z_k) = \sum_{r_1,\ldots,r_k \geqslant 0} N(r_1, \ldots, r_k) z_1^{r_1} \cdots z_k^{r_k} = \left(2 - \prod_{j=1}^{k}(1+z_j)\right)^{-1}. \tag{13.5}$$

We have $f(k,n) = N(n, \ldots, n)$, and so we need the diagonal terms of $F = F(z_1, \ldots, z_k)$. The function F is rational, so its singularity is small. Moreover, the singularities of F are difficult to visualize. However, in any singl̃e variable F is simple. We take advantage of this feature. Let

$$A(z) = \prod_{j=1}^{k-1}(1+z_j)\,, \tag{13.6}$$

where z stands for $(z_1, \ldots, z_{k-1}) \in \mathbb{C}^{k-1}$, and expand

$$\left(2 - \prod_{j=1}^{k}(1+z_j)\right)^{-1} = (2 - A(z)(1+z_k))^{-1} = \sum_{m=0}^{\infty} \frac{A(z)^m z_k^m}{(2 - A(z))^{m+1}} \, . \tag{13.7}$$

Therefore

$$N(r_1, \ldots, r_{k-1}, m) = \frac{1}{(2\pi i)^{k-1}} \int \cdots \int \frac{A(z)^m}{(2 - A(z))^{m+1}} \frac{dz_1}{z_1^{r_1+1}} \cdots \frac{dz_{k-1}}{z_{k-1}^{r_{k-1}+1}} \, . \tag{13.8}$$

The function whose coefficients we are trying to extract is now $A(z)^m/(2 - A(z))^{m+1}$, which is still rational. However, the interesting case for us is $m \to \infty$, which transforms the singularity into a large one. We are interested in the case $r_1 = r_2 = \cdots = r_{k-1} = r = n$. Then the integral in (13.8) can be shown to have a saddle point at $z_j = \rho$, $1 \leqslant j \leqslant k-1$, where $\rho = 2^{1/k} - 1$, and one obtains the estimate (Griggs et al. 1990)

$$f(k,n) = r^n n^{-(k-1)/2}\{(\rho\pi^{(k-1)/2}k^{1/2})^{-1}2^{(k^2-1)/(2k)} + O(n^{-1/2})\} \quad \text{as } n \to \infty \, . \tag{13.9}$$

$\boxtimes$

The examples above of applications of the multidimensional saddle point method all dealt with problems in a fixed dimension as various other parameters increase. A much more challenging problem is to apply this method when the dimension varies. A noteworthy case where this has been done successfully is the asymptotic enumeration of graphs with a given degree sequence by McKay and Wormald (1990).

Example 13.3 (*Simple labeled graphs of high degree*). Let $G(n; d_1, \ldots, d_n)$ be the number of labeled simple graphs on n vertices with degree sequence $d_1, d_2, \ldots, d_n$. Then $G(n; d_1, \ldots, d_n)$ is the coefficient of $z_1^{d_1} z_2^{d_2} \cdots z_n^{d_n}$ in

$$F = \prod_{\substack{j,k=1 \\ j<k}}^{n} (1 + z_j z_k) \, , \tag{13.10}$$

and so by Cauchy's theorem

$$G(n; d_1, \ldots, d_N) = (2\pi i)^{-n} \int \cdots \int F z_1^{-d_1-1} \cdots z_n^{-d_n-1} \, dz_1 \cdots dz_n \, , \tag{13.11}$$

where each integral is on a circle centered at the origin. Let all the radii be equal to some $r > 0$. The integrand takes on its maximum absolute value on the product of these circles at precisely the two points $z_1 = z_2 = \cdots = z_n = r$ and $z_1 = z_2 = \cdots = z_n = -r$. If $d_1 = d_2 = \cdots = d_n$, so that we consider only regular graphs, McKay and Wormald (1990) show that for an appropriate choice of the radius r, these two points are saddle points of the integrand, and succeed through careful analysis in proving that if dn is even, and $\min(d, n - d - 1) > cn(\log n)^{-1}$ for some $c > 2/3$, then

$$\begin{aligned} G(n, d, d, \ldots, d) = {} & 2^{1/2}(2\pi n\lambda^{d+1}(1 - \lambda)^{n-d})^{-n/2} \\ & \times \exp\left(\tfrac{-1+10\lambda-10\lambda^2}{12\lambda(1-\lambda)} + O(n^{-\zeta})\right) \end{aligned} \qquad (13.12)$$

as $n \to \infty$ for any $\zeta < \min(1/4, 1/2 - 1/(3c))$, where $\lambda = d/(n - 1)$.

McKay and Wormald (1990) also succeed in estimating the number of irregular graphs, provided that all the degrees d_j are close to a fixed d that satisfies conditions similar to those above. The proof is more challenging because different radii are used for different variables and the result is complicated to state. ☒

The McKay–Wormald estimate of Example 13.3 is a true tour de force. The problem is that the number of variables is n and so grows rapidly, whereas the integrand grows only like $\exp(cn^2)$ at its peak. More precisely, after transformations that remove obvious symmetries are applied the integrand near the saddle point drops off like $\exp(-n \sum \theta_j^2)$. This is just barely to allow the saddle point method to work, and the symmetries in the problem are exploited to push the estimates through. This approach can be applied to other problems (cf. McKay 1990), but it is hard to do. On the other hand, when the number of variables grows more slowly, multidimensional saddle-point contributions can be estimated without much trouble.

So far this section has been devoted primarily to multivariate functions with large singularities. However, there is also an extensive literature on small singularities. The main thread connecting most of these works is that of central and local limit theorems. Bender (1973) initiated this development in the setting of two-variable problems. We present some of his results, since they are simpler than the later and more general ones that will be mentioned at the end of this section.

Consider a double sequence of numbers $a_{n,k} \geqslant 0$. (Usually the $a_{n,k}$ are $\neq 0$ only for $0 \leqslant k \leqslant n$.) We will assume that

$$A_n = \sum_k a_{n,k} < \infty \qquad (13.13)$$

for all n, and define the normalized double sequence

$$p_n(k) = a_{n,k}/A_n\ . \qquad (13.14)$$

We will say that $a_{n,k}$ satisfies a central limit theorem if there exist functions σ_n and μ_n such that

$$\lim_{n\to\infty} \sup_x \left| \sum_{k \leqslant \sigma_n x + \mu_n} p_n(k) - (2\pi)^{-1/2} \int_{-\infty}^{x} \exp(-t^2/2)\,\mathrm{d}t \right| = 0 . \tag{13.15}$$

Equivalently, $p_n(k)$ is asymptotically normal with mean μ_n and variance σ_n^2.

Theorem 13.4 (Bender 1973). *Let $a_{n,k} \geqslant 0$, and set*

$$f(z, w) = \sum_{n,k \geqslant 0} a_{n,k} z^n w^k . \tag{13.16}$$

Suppose that there are (i) *a function $g(s)$ that is continuous and $\neq 0$ near $s = 0$,* (ii) *a function $r(s)$ with bounded third derivative near $s = 0$,* (iii) *an integer $m \geqslant 0$, and* (iv) *$\varepsilon, \delta > 0$ such that*

$$\left(1 - \frac{z}{r(s)}\right)^m f(z, \mathrm{e}^s) - \frac{g(z)}{1 - z/r(s)} \tag{13.17}$$

is analytic and bounded for

$$|z| < \varepsilon, \quad |z| < |r(0)| + \delta . \tag{13.18}$$

Let

$$\mu = -r'(0)/r(0), \quad \sigma^2 = \mu^2 - r''(0)/r(0) . \tag{13.19}$$

If $\sigma \neq 0$, then (13.15) *holds with $\mu_n = n\mu$ and $\sigma_n^2 = n\sigma^2$.*

A central limit theorem is useful, but it only gives information about the cumulative sums of the $a_{n,k}$. It is much better to have estimates for the individual $a_{n,k}$. We say that $p_n(k)$ (and $a_{n,k}$) satisfy a local limit theorem if

$$\lim_{n\to\infty} \sup_x \left| \sigma_n p_n(\lfloor \sigma_n x + \mu_n \rfloor) - (2\pi)^{-1/2} \exp(-x^2/2) \right| = 0 . \tag{13.20}$$

In general, we cannot derive (13.20) from (13.15) without some additional conditions on the $a_{n,k}$, such as unimodality (see Bender 1973). The other approach one can take is to derive (13.20) from conditions on the generating function $f(z, w)$.

Theorem 13.5 (Bender 1973). *Suppose that $a_{n,k} \geqslant 0$, and let $f(z, w)$ be defined by* (13.16). *Let $-\infty < a < b < \infty$. Define*

$$R(\varepsilon) = \{z \colon a \leqslant \operatorname{Re}(z) \leqslant b,\ |\operatorname{Im}(z)| \leqslant \varepsilon\} . \tag{13.21}$$

Suppose there exist $\varepsilon > 0$, $\delta > 0$, an integer $m \geqslant 0$, and function $g(s)$ and $r(s)$ such that

(i) *$g(s)$ is continuous and $\neq 0$ for $s \in R(\varepsilon)$,*
(ii) *$r(s) \neq 0$ and has a bounded third derivative for $s \in R(\varepsilon)$,*

(iii) *for* $s \in R(\varepsilon)$ *and* $|z| \leqslant |r(s)|(1+\delta)$, *the function defined by* (13.17) *is analytic and bounded,*

$$\text{(iv)} \quad \left(\frac{r'(\alpha)}{r(\alpha)}\right)^2 \neq \frac{r''(\alpha)}{r(\alpha)} \quad \textit{for } a \leqslant \alpha \leqslant b\,, \tag{13.22}$$

(v) $f(z, \mathrm{e}^s)$ *is analytic and bounded for*

$$|z| \leqslant |r(\mathrm{Re}(s))|(1+\delta) \quad \textit{and} \quad s \leqslant |\mathrm{Im}(s)| \leqslant \pi\,.$$

Then

$$a_{n,k} \sim \frac{n^m \mathrm{e}^{-\alpha k} g(\alpha)}{m!\, r(\alpha)^m \sigma_\alpha (2\pi)^{1/2}} \quad \textit{as } n \to \infty \tag{13.23}$$

uniformly for $a \leqslant \alpha \leqslant b$, *where*

$$\frac{k}{n} \doteq -\frac{r'(\alpha)}{r(\alpha)}\,, \tag{13.24}$$

$$\sigma_\alpha^2 = \left(\frac{k}{n}\right)^2 - \frac{r''(\alpha)}{r(\alpha)}\,. \tag{13.25}$$

There have been many further developments of central and local limit theorems for asymptotic enumeration since Bender's original work (1973). Currently the most powerful and general results are those of Gao and Richmond (1992). They apply to general multivariate problems, not only two-variable ones. Other papers that deal with central and local limit theorems or other multivariate problems with small singularities are Bender and Richmond (1983), Bender et al. (1983), Canfield (1977), Drmota (1994), Flajolet and Soria (1990, 1993), Gutjahr (1992), and Kirschenhofer (1987).

14. Mellin and other integral transforms

When the best generating function that one can obtain is an infinite sum, integral transforms can sometimes help. There is a large variety of integral transforms, such as those of Fourier and Laplace. The one that is most commonly used in asymptotic enumeration and analysis of algorithms is the Mellin transform, and it is the only one we will discuss extensively below. The other transforms do occur, though. For example, if $f(x) = \sum a_n x^n/n!$ is an exponential generating function of the sequence a_n, then the ordinary generating function of a_n can be derived from it using the Laplace transform

$$\begin{aligned} \int_0^\infty f(xy)\exp(-x)\,\mathrm{d}x &= \sum_n a_n y^n (n!)^{-1} \int_0^\infty x^n \exp(-x)\,\mathrm{d}x \\ &= \sum_n a_n y^n\,. \end{aligned} \tag{14.1}$$

(This assumes that the a_n are small enough to assure the integrals above converge and the interchange of summation and integration is valid.) Related integral transforms can be used to transform generating functions into other forms. For example, to transform an ordinary generating function $F(u) = \sum a_n u^n$ into an exponential one, we can use

$$\frac{1}{2\pi i} \int_{|u|=r} F(u) \exp(w/u)\, du \,. \tag{14.2}$$

The basic references for asymptotics of integral transforms are Davies (1978), Doetsch (1955), Oberhettinger (1974), and Sneddon (1972). This section will only highlight some of the main properties of Mellin transforms and illustrate how they are used. For a more detailed survey, especially to analysis of algorithms, see Flajolet et al. (1985).

Let $f(t)$ be a measurable function defined for real $t \geqslant 0$. The *Mellin transform* $f^*(z)$ of $f(t)$ is a function of the complex variable z defined by

$$f^*(z) = \int_0^\infty f(t) t^{z-1}\, dt \,. \tag{14.3}$$

If $f(t) = O(t^\alpha)$ as $t \to 0^+$ and $f(t) = O(t^\beta)$ as $t \to \infty$, then the integral in (14.3) converges and defines $f^*(z)$ to be an analytic function inside the "fundamental domain" $-\alpha < \mathrm{Re}(z) < -\beta$. As an example, for $f(t) = \exp(-t)$, we have $f^*(z) = \Gamma(z)$ and $\alpha = 0$, $\beta = -\infty$. There is an inversion formula for Mellin transforms which states that

$$f(t) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} f^*(z) t^{-z}\, dz \,, \tag{14.4}$$

and the integral is over the vertical line with $\mathrm{Re}(z) = c$. The inversion formula (14.4) is valid for $-\alpha < c < -\beta$, but much of its strength in applications comes from the ability to shift the contour of integration into wider domains to which $f^*(z)$ can be analytically continued.

The advantage of the Mellin transform is due largely to a simple property, namely that if $g(t) = af(bx)$ for b real, $b > 0$, then

$$g^*(z) = ab^{-z} f^*(z) \,. \tag{14.5}$$

This readily extends to show that if

$$F(t) = \sum_k \lambda_k f(\eta_k t) \tag{14.6}$$

(where the λ_k and $\eta_k > 0$ are such that the sum converges and $F(t)$ is well behaved), then

$$F^*(z) = \Big(\sum_k \lambda_k \eta_k^{-z}\Big) f^*(z) \,. \tag{14.7}$$

In particular, if

$$F(t) = \sum_{k=1}^{\infty} f(kt) , \tag{14.8}$$

then

$$F^*(z) = \left(\sum_{k=1}^{\infty} k^{-z}\right) f^*(z) = \zeta(z) f^*(z) , \tag{14.9}$$

where $\zeta(z)$ is the Riemann zeta function.

Example 14.1 (*Runs of heads in coin tosses*). What is R_n, the expected length of the longest run of heads in n tosses of a fair coin? Let $p(n,k)$ be the probability that there is no run of k heads in a coin tosses. Then

$$R_n = \sum_{k=1}^{n} k(p(n,k+1) - p(n,k)) . \tag{14.10}$$

We now apply the estimates of Example 9.3. To determine $p(n,k)$, we take $A = 00\cdots0$, and then $C_A(z) = z^{k-1} + z^{k-2} + \cdots + z + 1$, so $C_A(1/2) = 1 - 2^{-k}$. Hence (9.19) shows easily that in the important ranges where k is of order $\log n$, we have

$$p(n,k) \cong \exp(-n2^{-k}) , \tag{14.11}$$

and there R_n is approximated well by

$$r(n) = \sum_{k=0}^{\infty} k(\exp(-n2^{-k-1}) - \exp(-n2^{-k})) . \tag{14.12}$$

The function $r(t)$ is of the form (14.6) with

$$\lambda_k = k, \quad \eta_k = 2^{-k}, \quad f(t) = \exp(-t/2) - \exp(-t) , \tag{14.13}$$

is easily seen to be well behaved, and so for $-1 < \mathrm{Re}(z) < 0$,

$$r^*(z) = \left(\sum_{k=0}^{\infty} k2^{kz}\right) f^*(z) = 2^z(1-2^z)^{-2} f^*(z) . \tag{14.14}$$

Next, to determine $f^*(z)$, we note that for $\mathrm{Re}(z) > 0$ we have

$$\begin{aligned} f^*(z) &= \int_0^{\infty} f(t)t^{z-1}\,\mathrm{d}t = \int_0^{\infty} \mathrm{e}^{-t/2}t^{z-1}\,\mathrm{d}t - \int_0^{\infty} \mathrm{e}^{-t}t^{z-1}\,\mathrm{d}t \\ &= (2^z - 1)\Gamma(z) . \end{aligned} \tag{14.15}$$

By analytic continuation this relation holds for $-1 < \mathrm{Re}(z)$, and we find that for $-1 < \mathrm{Re}(z) < 0$,

$$r^*(z) = 2^z(2^z - 1)^{-1}\Gamma(z) . \tag{14.16}$$

We now apply the inversion formula to obtain

$$r(t) = \frac{1}{2\pi \mathrm{i}} \int_{-1/2-\mathrm{i}\infty}^{-1/2+\mathrm{i}\infty} 2^z(2^z-1)^{-1}\Gamma(z)t^{-z}\,\mathrm{d}z \,. \tag{14.17}$$

The integrand is a meromorphic function in the whole complex plane that drops off rapidly on any vertical line. We move the contour of integration to the line $\mathrm{Re}(z) = 1$. The new integral is $\mathrm{O}(t^{-1})$, and the residues at the poles (all on $\mathrm{Re}(z) = 0$) will give the main contribution to $r(t)$. There are first order poles at $z = 2\pi \mathrm{i} m \log 2$ for $m \in \mathbb{Z} \setminus \{0\}$ coming from $2^z = 1$, and a single second order pole at $z = 0$, since $\Gamma(z)$ has a first order pole there as well. A short computation of the residues gives

$$r(t) = \log_2 t - \sum_{h=-\infty}^{\infty} (\log 2)^{-1}\Gamma(-2\pi \mathrm{i}h(\log 2)^{-1})\exp(2\pi \mathrm{i}h \log_2 t) + \mathrm{O}(t^{-1}) \,. \tag{14.18}$$

⊠

There are other ways to obtain the same expansion (14.18) for $r(t)$ (cf. Guibas and Odlyzko 1980). The periodic oscillating component in $r(t)$ is common in problems involving recurrences over powers of 2. This happens, for example, in studies of register allocation and digital trees (Flajolet et al. 1979, Flajolet and Richmond 1992, Flajolet and Sedgewick 1986). The periodic function is almost always the same as the one in eq. (14.18), even when the combinatorics of the problem varies. Technically this is easy to explain, because of the closely related recurrences leading to similar Mellin transforms for the generating functions.

Mellin transforms are useful in dealing with problems that combine combinatorial and arithmetic aspects. For example, if $S(n)$ denotes the total number of 1's in the binary representations of $1, 2, \ldots, n-1$, then it was shown by Delange that

$$S(n) = \frac{1}{2} n \log_2 n + nu(\log_2 n) + \mathrm{o}(n) \quad \text{as } n \to \infty \,, \tag{14.19}$$

where $u(x)$ is a continuous, nowhere-differentiable function that satisfies $u(x) = u(x+1)$. The Fourier coefficients of $u(x)$ are known explicitly. Perhaps the best way to obtain these results is by using Mellin transforms. See Flajolet et al. (1994) and Stolarsky (1977) for further information and references.

Mellin transforms are often combined with other techniques. For example, sums of the form $s_n = \sum a_k \binom{n}{k}$ with oscillating a_k lead to generating functions

$$s(z) = \sum_k a_k w(z)^k \,. \tag{14.20}$$

The asymptotic behavior of $s(z)$ near its dominant singularity can sometimes be determined by applying Mellin transforms. For a detailed explanation of the approach, see Flajolet et al. (1985). Examples of the application of this technique can be found in Andrews (1976) and Meinardus (1954).

15. Functional equations, recurrences, and combinations of methods

Most asymptotic enumeration results are obtained from combinations of techniques presented in the previous sections. However, it is only rarely that the basic asymptotic techniques can be applied directly. This section describes a variety of methods and results that are not easy to categorize. They use combinations of methods that have been presented before, and sometimes develop them further. In most of the examples that will be presented, some relations for generating functions are available, but no simple closed-form formulas, and the problem is to deduce where the singularities lie and how the generating functions behave in their neighborhoods. Once that task is done, the previous methods can be applied to obtain asymptotics of the coefficients.

15.1. Implicit functions, graphical enumeration, and related topics

Example 15.1 (*Rooted unlabeled trees*). We sketch a proof that T_n, the number of rooted unlabeled trees with n vertices, satisfies the asymptotic relation (1.9). The functional equation (1.8) holds with $T(z)$ regarded as a formal power series. The first step is to show that $T(z)$ is analytic in a neighborhood of 0. This can be done by working exclusively with eq. (1.8). [There is an argument of this type in Harary and Palmer (1973, section 9.5).] Another way to prove analyticity of $T(z)$ is to use combinatorics to obtain crude upper bounds for T_n. We use a combination of these approaches. If a tree with $n \geqslant 2$ vertices has at least two subtrees at the root, we can decompose it into two trees, the first consisting of one subtree at the root, the other of the root and the remaining subtrees. This shows that

$$T_n \leqslant T_{n-1} + \sum_{k=1}^{n-1} T_k T_{n-k} , \quad n \geqslant 2 . \tag{15.1}$$

Therefore, if we define $a_1 = 1$, and

$$a_n = a_{n-1} + \sum_{k=1}^{n-1} a_k a_{n-k} , \quad n \geqslant 2 , \tag{15.2}$$

then we have $T_n \leqslant a_n$. Now if

$$A(z) = \sum_{n=1}^{\infty} a_n z^n ,$$

then the defining relation (15.2) yields the functional equation

$$A(z) - z = zA(z) + A(z)^2 , \tag{15.3}$$

so that

$$A(z) = (1 - z - (1 - 6z + z^2)^{1/2})/2 . \tag{15.4}$$

Since $A(z)$ is analytic in $|z| < 3 - 2\sqrt{2} = 0.171\,57\ldots$, we have

$$0 \leqslant T_n \leqslant a_n = \mathrm{O}(6^n) . \tag{15.5}$$

It will also be convenient to have an exponential lower bound for T_n. Let b_n be the number of rooted unlabeled trees in which every internal vertex has $\leqslant 2$ subtrees. Then $b_1 = 1$, $b_2 = 1$, and

$$b_n \geqslant \sum_{k=1}^{\lfloor (n-1)/2 \rfloor} b_k b_{n-k-1} \quad \text{for } n \geqslant 3 . \tag{15.6}$$

We use this to show that $b_n \geqslant (6/5)^n$ for $n \geqslant 7$. Direct computation establishes this lower bound for $7 \leqslant n \leqslant 14$, and for $n \geqslant 15$ we use induction and $b_n \geqslant b_k b_{n-k-1}$ with $k = \lfloor (n-1)/2 \rfloor$.

Since $T_n \geqslant b_n \geqslant (6/5)^n$, $T(z)$ converges only in $|z| < r$ for some r with $r < 1$. Since $T(0) = 0$, $|T(z)| \leqslant C_\delta |z|$ in $|z| \leqslant r - \delta$ for every $\delta > 0$, and therefore

$$u(z) = \sum_{k=2}^{\infty} T(z^k)/k \tag{15.7}$$

is analytic in $|z| < r^{1/2}$, and in particular at $z = r$. Therefore, although we know little about r and $u(z)$, we see that $T(z)$ satisfies $G(z, T(z)) = T(z)$, where

$$G(z, w) = z \exp(w + u(z)) \tag{15.8}$$

is analytic in z and w for all w and for $|z| < r^{1/2}$.

We will apply Theorem 10.13. First, though, we need to establish additional properties of $T(z)$. We have

$$T(z) \exp(-T(z)) = z \exp(u(z)) \to r \exp(u(r)) \quad \text{as } z \to r^- , \tag{15.9}$$

and $0 < r \exp(u(r)) < \infty$. Since $T(z)$ is positive and increasing for $0 < z < r$, $T(r)$, the limit of $T(z)$ as $z \to r^-$, must exist and be finite.

We next show that $T(r) = 1$. We have

$$\frac{\partial}{\partial w} G(z, w) = G(z, w) . \tag{15.10}$$

We know that $G(z, T(z)) = T(z)$ for $|z| < r$, and in particular for some z arbitrarily close to r. If $T(r) \neq 1$, then by (15.10)

$$\left. \frac{\partial}{\partial w} (G(z, w) - w) \right|_{w = T(z)} \neq 0 \tag{15.11}$$

in a neighborhood of $z = r$, and therefore $T(z)$ could be continued analytically to a neighborhood of $z = r$. This is impossible, since r is the radius of convergence of $T(z)$, and $T_n \geqslant 0$ implies by Theorem 10.4 that $T(z)$ has a singularity at $z = r$. Therefore we must have $T(r) = 1$, and $G_w(r, T(r)) = 1$.

We have now shown that conditions (i) and (ii) of Theorem 10.13 hold with the r of that theorem the same as the r we have defined and $s = T(r) = 1$, $\delta = r^{1/2} - r$. Condition (iii) is easy to verify. Finally, the conditions on the coefficients of $T(z)$ and $G(z, w)$ are clearly satisfied.

Since Theorem 10.13 applies, we do obtain an asymptotic expansion for T_n of the form (1.9), with C given by the formula (10.64). It still remains to determine r and C. No closed-form expressions are known for these constants. They are conjectured to be transcendental and algebraically independent of standard constants such as π and e, but no proof is available. Numerically, however, they are simple to compute. Note that

$$G_z(r,1) = \exp(1 + u(r))(1 + ru'(r)) = r^{-1} + u'(r) \ , \tag{15.12}$$

$$G_{ww}(r,1) = 1 \ , \tag{15.13}$$

so we only need to compute r and $u'(r)$. These quantities can be computed along with $u(r)$ in the same procedure. The basic numerical procedure is to determine r as the positive solution to $T(r) = 1$. To determine $T(x)$ for any positive x, we take any approximation to the $T(x^k)$, $k \geqslant 1$ (starting initially with x^k as an approximation to $T(x^k)$, say), and combine it with (1.8) (applied with $z = x^m$, $m \geqslant 1$) to obtain improved approximations. This procedure can be made rigorous. Upper bounds for r, $u(r)$, and $u'(r)$ are especially easy. Since $T_1 = 1$, $T(x) \geqslant x$ for $0 < x < 1$, and therefore, $T(x^k) \geqslant x^k$ for $k \geqslant 1$. Suppose that we start with a fixed value of x and derive some lower bounds of the form $T(x^k) \geqslant u_k^{(1)} \geqslant 0$ for $k \geqslant 1$. Then the functional equation (1.8) implies

$$T(x^m) \geqslant u_m^{(2)} = x \exp\left(\sum_{k=1}^{\infty} u_{km}/k\right), \qquad m \geqslant 1 \ . \tag{15.14}$$

This process can be iterated several more times, and to keep the computation manageable, we can always set $u_k^{(j)} = 0$ for $k \geqslant k_0$. If we ever find a lower bound $T(x) > 1$ by this process, then we know that $r < x$, since $T(r) = 1$. Lower bounds for r are slightly more complicated. ⊠

We mention here that if U_n denotes the number of unlabeled trees, then the ordinary generating function $U(z) = \sum U_n z^n$ satisfies

$$U(z) = T(z) - T(z)^2/2 + T(z^2)/2 \ . \tag{15.15}$$

Using the results from Example 15.1 about the analytic behavior of $T(z)$, it can be shown that

$$U_n \sim C' r^{-n} n^{-5/2} \ , \tag{15.16}$$

where $r = 0.338\,321\,9\ldots$ is the same as before, while $C' = 0.534\,948\,5\ldots$

Example 15.2 (*Leftist trees*). Let a_n denote the number of leftist trees of size n [i.e., rooted planar trees with n leaves, such that in any subtree S, the leaf nearest to the root of S is in the right subtree of S (Knuth 1973b)]. Then $a_1 = a_2 = a_3 = 1$, $a_4 = 2$, $a_5 = 4$. No explicit formula for a_n is known. Even the recurrences for the a_n are complicated, and involve auxiliary sequences. If

$$f(z) = \sum_{n=1}^{\infty} a_n z^n \tag{15.17}$$

denotes the ordinary generating function of a_n, then the combinatorially derived recurrences for the a_n show that (Kemp 1987)

$$f(z) = z + \frac{1}{2} f(z)^2 + \frac{1}{2} \sum_{m=1}^{\infty} g_m(z)^2 , \tag{15.18}$$

where the auxiliary generating functions $g_m(z)$ (which enumerate leftist trees with the leftmost leaf at distance $m - 1$ from the root) satisfy

$$g_1(z) = z, \quad g_2(z) = zf(z), \quad g_{m+1}(z) = g_m(z) \left[f(z) - \sum_{j=1}^{m-1} g_j(z) \right] , \quad m \geqslant 2 , \tag{15.19}$$

and

$$f(z) = \sum_{m=1}^{\infty} g_m(z) . \tag{15.20}$$

These generating-function relations might not seem promising. If r is the smallest singularity of $f(z)$, then $\sum g_m(z)^2$ is not analytic at r, so we cannot apply Theorem 10.13 in the way it was used in Example 15.1. However, Kemp (1987) has sketched a proof that the analytic behavior of $f(z)$ is of the same type as that involved in functions covered by Theorem 10.13, so that it has a dominant square-root singularity, and therefore

$$a_n = \alpha c^n n^{-3/2} + O(c^n n^{-5/2}) , \tag{15.21}$$

where

$$\alpha = 0.250\,363\,429 \ldots, \quad c = 2.749\,487\,902 \ldots . \tag{15.22}$$

The constants α and c are not known explicitly in terms of other standard numbers such as π or e, but they can be computed efficiently. The $\alpha c^n n^{-3/2}$ term in (15.21) gives an approximation to a_n that is accurate to within 4% for $n = 10$, and within 0.4% for $n = 100$. Thus asymptotic methods yield an approximation to a_n which is factory for many applications. Further results about leftist trees can be found ›mp (1990). ⊠

15.2. Nonlinear iteration and tree parameters

Example 15.3 (*Heights of binary trees*). A binary tree (Knuth 1973b) is a rooted tree with unlabeled nodes, in which each node has 0 or 2 successors, and left and right successors are distinguished. The size of a binary tree is the number of internal nodes, i.e., the number of nodes with two successors. We let B_n denote the number of binary trees of size n, so that $B_0 = 1$ (by convention), $B_1 = 1$, $B_2 = 2$, $B_3 = 5, \ldots$. Let

$$B(z) = \sum_{n=0}^{\infty} B_n z^n . \tag{15.23}$$

Since each nonempty binary tree consists of the root and two binary trees (the left and right subtrees), we obtain the functional equation

$$B(z) = 1 + zB(z)^2 . \tag{15.24}$$

This implies that

$$B(z) = \frac{1 - (1 - 4z)^{1/2}}{2z} , \tag{15.25}$$

so that

$$B_n = \frac{1}{n+1}\binom{2n}{n} , \tag{15.26}$$

and the B_n are the Catalan numbers. The formula (4.4) [easily derivable from Stirling's formula (4.1)] shows that

$$B_n \sim \pi^{-1/2} n^{-3/2} 4^n \quad \text{as } n \to \infty . \tag{15.27}$$

The height of a binary tree is the number of nodes along the longest path from the root to a leaf. The distribution of heights in binary trees of a given size does not have exact formulas like that of (15.26) for the number of binary trees of a given size. There are several problems on heights that have been answered only asymptotically, and with varying degrees of success. The most versatile approach is through recurrences on generating functions. Let $B_{h,n}$ be the number of binary trees of size n and height $\leqslant h$, and let

$$b_h(z) = \sum_{n=0}^{\infty} B_{h,n} z^n . \tag{15.28}$$

Then

$$b_0(z) = 0, \quad b_1(z) = 1 , \tag{15.29}$$

and an extension of the argument that led to the relation (15.24) yields

$$b_{h+1}(z) = 1 + zb_h(z)^2 , \quad h \geqslant 0 . \tag{15.30}$$

The $b_h(z)$ are polynomials in z of degree $2^{h-1}-1$ for $h \geqslant 1$. Unfortunately there is no simple formula for them like eq. (15.25) for $B(z)$, and one has to work with the recurrence (15.30) to obtain many of the results about heights of binary trees. Different problems involve study of the recurrence in different ranges of values of z, and the behavior of the recurrence varies drastically.

For any fixed z with $|z| \leqslant 1/4$, $b_h(z) \to B(z)$ as $h \to \infty$. For $|z| > 1/4$ the behavior of $b_h(z)$ is more complicated, and is a subject of of nonlinear dynamics (Devaney 1989). (It is closely related to the study of the Mandelbrot set.) For any real z with $z > 1/4$, $b_h(z) \to \infty$ as $h \to \infty$. To study the distribution of the $B_{h,n}$ as n varies for h fixed, but large, it is necessary to investigate this range of rapid growth. It can be shown (Flajolet and Odlyzko 1984) that for any λ_1 and λ_2 with $0 < \lambda_1 < \lambda_2 < 1/2$,

$$B_{h,n} = \frac{\exp(2^{h-1}(\beta(r) - r\beta'(r)\log r))}{2^{(h-1)/2}(2\pi(r^2\beta''(r) + r\beta'(r)))^{1/2}}(1 + \mathrm{O}(2^{-h/2})) \tag{15.31}$$

uniformly as $h, n \to \infty$ with

$$\lambda_1 < n/2^h < \lambda_2 , \tag{15.32}$$

where the function $\beta(x)$ is defined for $1/4 < x < \infty$ by

$$\beta(x) = \log x + \sum_{j=1}^{\infty} 2^{-j} \log\left(1 + \frac{1}{b_j(x) - 1}\right) , \tag{15.33}$$

and r is the unique solution in $(1/4, \infty)$ to

$$r\beta'(r) = n2^{-h+1} . \tag{15.34}$$

The formula (15.31) might appear circular, in that it describes the behavior of the coefficients $\beta_{h,n}$ of the polynomial $b_h(z)$ in terms of the function $\beta(z)$, which is defined by $b_h(z)$ and all the other $b_j(z)$. However, the series (15.33) for $\beta(z)$ converges rapidly, so that only the first few of the $b_h(z)$ matter in obtaining approximate answers, and computation using (15.33) is efficient. The function $\beta(z)$ is analytic in a region containing the real half-line $x > 1/4$, so the behavior of the $B_{h,n}$ is smooth. It is also known (Flajolet and Odlyzko 1984) that the behavior of $B_{h,n}$ as a function of n is Gaussian near the peak, which occurs at $n \sim 2^{h-1} \cdot 0.628\,968\ldots$. The distribution of $B_{h,n}$ is not Gaussian throughout the range (15.32), though.

The proof of the estimate (15.31) is derived from the estimate

$$b_h(z) = \exp(2^{h-1}\beta(z) - \log z)(1 + \mathrm{O}(\exp(-\varepsilon 2^h))) , \tag{15.35}$$

valid in a region along the half-axis $x > 1/4$. The estimates for the coefficients $B_{h,n}$ are obtained by applying the saddle point method. Because of the doubly-exponential rate of growth of $b_h(z)$ for z close to the real axis, it is easy to show that on the circle of integration, the region away from the real axis contributes a negligible amount to $B_{h,n}$. The relation (15.35) is sufficient, together with the smoothness properties of $\beta(z)$, to estimate the contribution of the integral near the real axis. To prove (15.35), one proceeds as in Example 9.8. However, greater care is required because of the complex variables that occur and the need for estimates that are uniform in the variables. The basic recurrence (15.30) shows that

$$\begin{aligned}\log b_{h+1}(z) &= 2\log b_h(z) + \log z + \log\left(1 + \frac{1}{zb_h(z)^2}\right)\\ &= 2\log b_h(z) + \log z + \log\left(1 + \frac{1}{b_{h+1}(z) - 1}\right) .\end{aligned} \tag{15.36}$$

Iterating this relation, we find that for $h \geqslant 1$,

$$\begin{aligned}\log b_{h+1}(z) &= 2^{h+1}\log b_1(z) + (2^h - 1)\log z + \sum_{k=0}^{h-1} 2^k \log\left(1 + \frac{1}{b_{h+1-k}(z) - 1}\right)\\ &= 2^h\left\{\log z + \sum_{j=1}^{h+1} 2^{-j}\log\left(1 + \frac{1}{b_j(z) - 1}\right)\right\} - \log z .\end{aligned} \tag{15.37}$$

The basic equation (15.35) then follows. The technical difficulty is in establishing rigorous bounds for the error terms in the approximations. Details are presented in Flajolet and Odlyzko (1984).

Most of the binary trees of a given height h are large, with about $0.3 \cdot 2^h$ internal nodes. This might give the misleading impression that most binary trees are close to the full binary tree of a similar size. However, if we consider all binary trees of a given size n, the average height is on the order of $n^{1/2}$, so that they are far from the full balanced binary trees. The methods that are used to study the average height are different from those used for trees of a fixed height. The basic approach of Flajolet and Odlyzko (1984) is to let

$$H_n = \sum_{\substack{T\\ |T|=n}} \mathrm{ht}(T) ,$$

where the sum is over the binary trees T of size n, and $\mathrm{ht}(T)$ is the height of T. Then the average height is just H_n/B_n.

The generating function for the H_n is

$$H(z) = \sum_{n=0}^{\infty} H_n z^n = \sum_{h\geqslant 0}(B(z) - b_h(z)) , \tag{15.38}$$

and the analysis of Flajolet and Odlyzko (1984) proceeds by investigating the behavior of $H(z)$ in a wedge-shaped region of the type encountered in section 11.1.

If we let

$$\varepsilon(z) = (1-4z)^{1/2} , \tag{15.39}$$

$$e_h(z) = (B(z) - b_h(z))/(2B(z)) , \tag{15.40}$$

then the recurrence (15.30) yields

$$e_{h+1}(z) = (1-\varepsilon(z))e_h(z)(1-e_h(z)) , \quad e_0(z) = 1/2 . \tag{15.41}$$

Extensive analysis of this relation yields an approximation to $e_h(z)$ of the form

$$e_h(z) \approx \frac{\varepsilon(z)(1-\varepsilon(z))^h}{1-(1-\varepsilon(z))^h} , \tag{15.42}$$

valid for $|\varepsilon(z)|$ sufficiently small, $|\text{Arg } \varepsilon(z)| < \pi/4 + \delta$ for a fixed $\delta > 0$. [The precise error terms in this approximation are complicated, and are given in Flajolet and Odlyzko (1984).] This then leads to an expansion for $H(z)$ in a sector $|z - 1/4| < \alpha$, $\pi/2 - \beta < |\text{Arg}(z - 1/4)| < \pi/2 + \beta$ of the form

$$H(z) = -2\log(1-4z) + K + \mathrm{O}(|1-4z|^v) , \tag{15.43}$$

where v is any constant, $v < 1/4$, and K is a fixed constant. Transfer theorems of section 11.1 now yield the asymptotic estimate

$$H_n \sim 2n^{-1}4^n \quad \text{as } n \to \infty . \tag{15.44}$$

When we combine (15.44) with (15.27), we obtain the desired result that the average height of a binary tree of size n is $\sim 2(\pi n)^{1/2}$ as $n \to \infty$.

Distribution results about heights of binary trees can be obtained by investigating the generating functions

$$\sum_{h \geqslant 0} h^r(B(z) - b_h(z)) . \tag{15.45}$$

This procedure, carried out in Flajolet and Odlyzko (1984) by using modifications of the approach sketched above for the average height, obtains asymptotics of the moments of heights. The method mentioned in section 6.5 then leads to a determination of the distribution. However, the resulting estimates do not say much about heights far away from the mean. A more careful analysis of the behavior of $e_h(z)$ can be used Flajolet et al. (1993a) to show that if $x = h/(2n^{1/2})$, then

$$\frac{B_{h,n} - B_{h-1,n}}{B_n} \sim 2xn^{-1/2} \sum_{m=1}^{\infty} m^2(2m^2x^2 - 3)\,\mathrm{e}^{-m^2x^2} \tag{15.46}$$

as $n, h \to \infty$, uniformly for $x = \mathrm{o}((\log n)^{1/2})$, $x^{-1} = \mathrm{o}((\log n)^{1/2})$.

For extremely small and large heights, different methods are used. It follows from Flajolet et al. (1993a) that

$$\frac{B_{h,n} - B_{h-1,n}}{B_n} \leqslant \exp(-c(h^2/n + n/h^2)) \tag{15.47}$$

for a constant $c > 0$, which shows that extreme heights are infrequent. [The estimates in Flajolet et al. (1993a) are more precise than (15.47).] Bounds of the above form for small heights are obtained in Flajolet et al. (1993a) by studying the behavior of the $b_h(z)$ almost on the boundary between convergence and divergence, using the methods of Wright et al. (1986). Let x_h be the unique positive root of $b_h(z) = 2$. Note that $B(1/4) = 2$, and each coefficient of the $b_h(z)$ is nondecreasing as $h \to \infty$. Therefore $x_2 > x_3 > \cdots > 1/4$. More effort shows (Flajolet et al. 1993a) that x_h is approximately $1/4 + \alpha h^{-2}$ for a certain $\alpha > 0$. This leads to an upper bound for $B_{h,n}$ by Lemma 8.4. Bounds for trees of large heights are even easier to obtain, since they only involve upper bounds for the $b_h(z) - b_{h-1}(z)$ inside the disk of convergence $|z| < 1/4$. ☒

In addition to the methods of Flajolet and Odlyzko (1982, 1984) and Flajolet et al. (1993a) that were mentioned above, there are also other techniques for studying heights of trees, such as those of Brown and Shubert (1984) and Rényi and Szekeres (1967). However, there are problems about obtaining fully rigorous proofs that way. [See the remarks in Flajolet et al. (1993a) on this topic.] Most of these methods can be extended to study related problems, such as those of diameters of trees (Szekeres 1982).

The results of Example 15.3 can be extended to other families of trees (cf. Flajolet and Odlyzko 1982, 1984, Flajolet et al. 1993a). What matters in obtaining results such as those of the above example are the form of the recurrences, and especially the positivity of the coefficients.

Example 15.4 (*Enumeration of* 2,3-*trees* (Odlyzko 1982)). Height-balanced trees satisfy different functional equations than unrestricted trees, which results in different analytic behavior of the generating functions, and different asymptotics. Consider 2, 3-trees; i.e., rooted, oriented trees such that each nonleaf node has either two or three successors, and in which all root-to-leaf paths have the same length. If a_n is the number of 2, 3-trees with exactly n leaves, then $a_1 = a_2 = a_3 = a_4 = 1$, $a_5 = 2, \ldots$, and the generating function

$$f(z) = \sum_{n=1}^{\infty} a_n z^n \tag{15.48}$$

satisfies the functional equation

$$f(z) = z + f(z^2 + z^3) . \tag{15.49}$$

Iteration of the recurrence (15.49) leads to

$$f(z) = \sum_{k=0}^{\infty} Q_k(z) , \tag{15.50}$$

where $Q_0(z) = z$, $Q_{k+1}(z) = Q_k(z^2 + z^3)$, provided the series (15.50) converges. The Taylor series (15.48) converges only in $|z| < \phi^{-1}$, where $\phi = (1 + 5^{1/2})/2$ is the "golden ratio". Study of the polynomials $Q_k(z)$ shows that the expansion (15.50) converges in a region

$$D = \{z\colon |z| < \phi^{-1} + \delta,\ |\mathrm{Arg}(z - \phi^{-1})| > \pi/2 - \varepsilon\} \tag{15.51}$$

for certain δ, $\varepsilon > 0$, and that inside D,

$$f(z) = -c\log(\phi^{-1} - z) + w(\log(\phi^{-1} - z)) + \mathrm{O}(|\phi^{-1} - z|)\ , \tag{15.52}$$

where $c = [\phi \log(4 - \phi)]^{-1}$, and $w(t)$ is a nonconstant function, analytic in a strip $|\mathrm{Im}(t)| < \eta$ for some $\eta > 0$, such that $w(t + \log(4 - \phi)) = w(t)$. The expression (15.52) only has to be proved in a small vicinity of ϕ^{-1} (intersected with D, of course). Since

$$Q(\phi^{-1} + \nu) = \phi^{-1} + (4 - \phi)\nu + \mathrm{O}(|\nu|^2) \tag{15.53}$$

(so that ϕ^{-1} is a repelling fixed point of Q), behavior like that of (15.52) is to be expected, and with additional work can be rigorously shown to hold. Once the expansion (15.52) is established, singularity analysis techniques can then be applied to deduce that

$$a_n \sim \frac{\phi^n}{n} u(\log n) \quad \text{as } n \to \infty\ , \tag{15.54}$$

where $u(t)$ is a positive nonconstant continuous function that satisfies $u(t) = u(t + \log(4 - \phi))$, and has mean value $(\phi \log(4 - \phi))^{-1}$. For details, see Odlyzko (1982).

The same methods can be applied to related families of trees, such as those of B-trees. ☒

The results of Example 15.3 and the generalizations mentioned above all apply only to the standard counting models, in which all trees with a fixed value of some simple property, such as size or height, are equally likely. Often, especially in computer-science applications, it is necessary to study trees produced by some algorithm, and consider all outputs of this algorithm as equally likely. For example, in sorting it is natural to consider all permutations of n elements as equally probable. If random permutations are used to construct binary search trees, then the distribution of heights will be different from that in the standard model, and the two trees of maximal height will have probability of $2/n!$ of occurring. The average height turns out to be $\sim c \log n$ as $n \to \infty$, for $c = 4.311\ldots$, a certain constant given as a solution to a transcendental equation. This was shown by Devroye (1986) (see also Devroye 1987) by an application of the theory of branching processes. For a detailed exposition of this method and other applications to similar problems, see Mahmoud (1992). The basic generating-function approach that we have used in most of this chapter leads to functional iterations which have not been solved so far.

15.3. Differential and integral equations

Section 9.2 showed that differential equations arise naturally in analyzing linear recurrences of finite order with rational coefficients. There are other settings where they arise even more naturally. As is true of nonlinear iterations in the previous section and the functional equations of the next one, differential and integral equations are typically used to extract information about singularities of generating functions. We have already seen in Example 9.4 and other cases that differential equations can yield an explicit formula for the generating function, from which it is easy to deduce what the singularities are and how they affect the asymptotics of the coefficients. Most differential equations do not have a closed-form solution. However, it is often still possible to derive the necessary information about analytic behavior even when there is no explicit formula for the solution. We demonstrate this with a brief sketch of a recent analysis of this type (Flajolet and Lafforgue 1994). Other examples can be found in Mahmoud (1992).

Example 15.5 (*Search costs in quadtrees* (Flajolet and Lafforgue 1994)). Quadtrees are a well-known data structure for multidimensional data storage Gonnet and Baeza-Yates 1991). Consider a d-dimensional data space, and let n points be drawn independently from the uniform distribution in the d-dimensional unit cube. We take d fixed and $n \to \infty$. Suppose that the first $n-1$ points have already been inserted into the quadtree, and let D_n be the search cost (defined as the number of internal nodes traversed) in inserting the nth item. The result of Flajolet and Lafforgue (1994) is that D_n converges in distribution to a Gaussian law when $n \to \infty$. If μ_n and σ_n denote the mean and standard deviation of D_n, respectively, then

$$\mu_n \sim 2d^{-1}\log n, \quad \sigma_n \sim d^{-1}(2\log n)^{1/2} \quad \text{as } n \to \infty\,, \tag{15.55}$$

and for all real $\alpha < \beta$, as $n \to \infty$,

$$\Pr(\alpha\sigma_N < D_n - \mu_n < \beta\sigma_n) \sim (2\pi)^{-1/2}\int_\alpha^\beta \exp(-x^2/2)\,\mathrm{d}x\,. \tag{15.56}$$

The results for μ_n and σ_n had been known before, and required much simpler techniques for their solution, see Mahmoud (1992). It was only necessary to study asymptotics of ordinary differential equations in a single variable. To obtain distribution results for search costs, it was necessary to study bivariate generating functions. The basic relation is

$$\sum_k \Pr\{D_n = k\}u^k = (2^d u - 1)^{-1}(\phi_n(u) - \phi_{n-1}(u))\,, \tag{15.57}$$

where the polynomials $\phi_n(u)$ have the bivariate generating function

$$\Phi(u,z) = \sum_{n=0}^{\infty} \phi_n(u) z^n \tag{15.58}$$

which satisfies the integral equation

$$\begin{aligned}\Phi(u,z) = 1 + 2^d u \int_0^z \frac{\mathrm{d}x_1}{x_1(1-x_1)} \int_0^{x_1} \frac{\mathrm{d}x_2}{x_2(1-x_2)} \int_0^{x_2} \frac{\mathrm{d}x_3}{x_3(1-x_3)} \cdots \\ \int_0^{x_{d-2}} \frac{\mathrm{d}x_{d-1}}{x_{d-1}(1-x_{d-1})} \int_0^{x_{d-1}} \Phi(u,x_d)\frac{\mathrm{d}x_d}{1-x_d}\,.\end{aligned} \tag{15.59}$$

This integral equation can easily be reduced to an equivalent differential equation, which is what is used in the analysis. For $d = 1$ there is an explicit solution

$$\Phi(u,z) = (1-z)^{-2u} , \tag{15.60}$$

which shows that D_n can be expressed in terms of Stirling numbers. This is not surprising, since for $d = 1$ the quadtree reduces to the binary search tree, for which these results were known before. For $d = 2$, $\Phi(u,z)$ can be expressed in terms of standard hypergeometric functions. However, for $d \geqslant 3$ there do not seem to be any explicit representations of $\Phi(u,z)$. Flajolet and Lafforgue use a singularity perturbation method to study the behavior of $\Phi(u,z)$. They start out with the differential system derivable in standard way from the differential equation associated to (15.59) (i.e., a system of d linear differential equations in z with coefficients that are rational in z). Since only values of u close to 1 are important for the distribution results, they regard u as a perturbation parameter of this system. For every fixed u, they determine the dominant singularity of the linear differential system in the variable z, using the indicial equations that are standard in this setting. It turns out that the dominant singularity is a regular one at $z = 1$, and

$$\Phi(u,z) \approx c(u)(1-z)^{-2u^{1/d}} , \tag{15.61}$$

at least for z and u close to 1. This behavior of $\Phi(u,z)$ is then used (in its more precise form, with explicit error terms) to deduce, through the transfer theorem methods explained in section 11, the behavior of $\phi_n(u)$:

$$\phi_n(u) \approx c(u)\Gamma(2u^{1/d})^{-1}n^{2u^{1/d}-1} . \tag{15.62}$$

This form, again in a more precise formulation, is then used to deduce that the behavior of D_n is normal near its peak, and that the tails of the distribution are small. ⊠

15.4. Functional equations

One area that needs and undoubtedly will receive much more attention is that of complicated nonlinear relations for generating functions. Even in a single variable our knowledge is limited. Some of the work of Mahler (1976, 1981, 1983), devoted to functions $f(z)$ satisfying equations of the form $p(f(z), f(z^g)) = 0$, where $p(u,v)$ is a polynomial, shows that it is possible to extract information about the analytic behavior of $f(z)$ near its singularities. This can then be used to study the coefficients.

Sometimes seemingly complicated functional equations do have easy solutions.

Example 15.6 (*A pebbling game*). In a certain pebbling game (Chung et al. 1995), minimal configurations of size n are counted by $T_n(0)$, where $T_n(x)$ is a polynomial that satisfies $T_n(x) = 0$ for $0 \leqslant n \leqslant 2$, $T_3(x) = 4x + 2x^2$, and for $n \geqslant 3$,

$$T_{n+1}(x) = x^{-1}(1+x)^2T_n(x) - x^{-1}T_n(0) + xT_n'(0) . \tag{15.63}$$

The coefficients of $T_n(x)$ are $\geqslant 0$, and

$$T_{n+1}(1) \leqslant 4T_n(1) + T_n(1) + 1 \leqslant 6T_n(1)\ , \tag{15.64}$$

so clearly each coefficient of $T_n(x)$ is $\leqslant 6^n$, say. Let

$$f(x,y) = \sum_{n=0}^{\infty} T_n(x)y^n\ . \tag{15.65}$$

The bound on $T_n(1)$ shows that $f(x,y)$ is analytic in x and y for $|x| < 1$, $|y| < 1/6$, say, with x and y complex. Then the recurrence (15.63) leads to the functional equation

$$(x - y(1+x)^2)f(x,y) = 2x^2(2+x)y^3 - yf(0,y) + x^2yf_x(0,y)\ , \tag{15.66}$$

where $f_x(x,y)$ is the partial derivative of $f(x,y)$ with respect to x. We now differentiate eq. (15.66) with respect to x and set $x = 0$. We find that

$$(1-2y)f(0,y) = yf_x(0,y)\ , \tag{15.67}$$

and therefore

$$(x - y(1+x)^2)f(x,y) = 2x^2(2+x)y^3 - [y + (2y-1)x^2]f(0,y)\ . \tag{15.68}$$

When

$$x = y(1+x)^2\ , \tag{15.69}$$

the left side of eq. (15.68) vanishes, and eq. (15.68) yields the value of $f(0,y)$. Now eq. (15.69) holds for

$$x = (2y)^{-1}(1 - 2y \pm (1-4y)^{1/2})\ .$$

To ensure that (15.69) holds for x and y both in a neighborhood of 0, we set

$$g(y) = (2y)^{-1}(1 - 2y - (1-4y)^{1/2})\ . \tag{15.70}$$

Then $g(y) = y(1+g(y))^2$, $g(y)$ is analytic for $|y|$ small, and so substituting $x = g(y)$ in eq. (15.68) yields

$$\begin{aligned}&(1 - 7y + 14y^2 - 9y^3)f(0,y)\\ &\quad = y^2((1-4y)^{1/2}(1 - 3y + y^2) - 1 + 5y - y^2 - 6y^3)\ .\end{aligned} \tag{15.71}$$

Thus $f(0,y)$ is an algebraic function of y. Equation (15.71) was proved only for $|y|$ small, but it can now be used to continue $f(0,y)$ analytically to the entire complex plane with the exception of a slit from $1/4$ to infinity along the positive real axis. There is a first order pole at $y = 1/r$, with $r = 4.147\,899\,035\,7\ldots$, the positive root of

$$r^3 - 7r^2 + 14r - 9 = 0\ , \tag{15.72}$$

and no other singularities in $|y| < 1/4$. Hence we obtain

$$T_n(0) = [y^n]f(0,y) = cr^n + O((4+\varepsilon)^n) \tag{15.73}$$

as $n \to \infty$, for every $\varepsilon > 0$, where c is an algebraic number that can be given explicitly in terms of r.

The value of $f(0,y)$ is determined by eq. (15.71), and together with eq. (15.68) gives $f(x,y)$ explicitly as an algebraic function of x and y. The resulting expression can then be used to determine other coefficients of the polynomials $T_n(x)$. ☒

Example 15.6 was easy to present because of the special structure of the functional equation. The main trick was to work on the variety defined by eq. (15.69), on which the main term vanishes, so that one can analyze the remaining terms. The same basic approach also works in more complicated situations. The analysis of certain double queue systems leads to two-variable generating functions for the equilibrium probabilities that satisfy equations such as the following one, obtained by specializing the problem treated in Flatto and Hahn (1984):

$$Q(z,w)f(z,w) = 2z(w-1)f(z,0) + 3w(z-1)f(0,w) \ , \tag{15.74}$$

valid for complex z and w with $|z|, |w| \leqslant 1$, where

$$Q(z,w) = 6zw - 3w - 2z - z^2w^2 \ . \tag{15.75}$$

The generating function $f(z,w)$ is analytic in z and w. What makes this problem tractable is that on the algebraic curve in two-dimensional complex space defined by $Q(z,w) = 0$, the quantity on the right-hand side of eq. (15.74) has to vanish, and this imposes stringent conditions on $f(z,0)$ and $f(0,w)$, which leads to their determination. Once $f(z,0)$ and $f(0,w)$ are found, $f(z,w)$ is defined by eq. (15.74), and one can determine the asymptotics of its coefficients. Treatment of functional equations of the type (15.74) was started by Malyshev (1972). For recent work and references to other papers in this area, see Flatto (1989) and Flatto and Hahn (1984). This approach has so far been successful only for two-variable problems with $Q(z,w)$ of low degree. Moreover, the mathematics of the solution is far deeper than that used in Example 15.6.

16. Other methods

This section mentions a variety of methods that are not covered elsewhere in this chapter but are useful in asymptotic enumeration. Most are discussed briefly, since they belong to large and well-developed fields that are beyond the scope of this survey.

16.1. Permanents

Van der Waerden's conjecture, proved by Falikman (1981) and Egorychev (1981), can be used to obtain lower bounds for certain enumeration problems. It states that if A is an $n \times n$ matrix that is doubly stochastic (entries $\geqslant 0$, all row and column sums equal to 1) then the permanent of A satisfies $\operatorname{per}(A) \geqslant n^{-n}n!$. [For most asymptotic problems it is sufficient to rely on an earlier result of Bang (1976) and Friedland (1979) which gives a lower bound of $\operatorname{per}(A) \geqslant \mathrm{e}^{-n}$ that is worse only by a factor of $n^{1/2}$.] There is also an upper bound for permanents. Minc's conjecture, proved first by Bragman and in a simpler way by Schrijver (1978) states that an $n \times n$ matrix A with 0, 1 entries and row sums $r_1, \ldots, r_n$ has

$$\operatorname{per}(A) \leqslant \prod_{j=1}^{n} (r_j!)^{1/r_j} .$$

We now show how these results can be applied.

Example 16.1 (*Latin rectangles*). Suppose we are given a $k \times n$ Latin rectangle, $k < n$, so that the symbols are $1, 2, \ldots, n$, and no symbol appears twice in any row or column. In how many ways can we extend this rectangle to a $(k+1) \times n$ Latin rectangle? To get a lower bound, form an $n \times n$ matrix $B = (b_{ij})$, with $b_{ij} = 1$ if i does not appear in column j of the rectangle, and $b_{ij} = 0$ otherwise. Then the row and column sums of B are all equal to $n-k$, so $(n-k)^{-1}B$ is doubly stochastic. Therefore $\operatorname{per}(B)$, which equals the desired number of ways of extending the rectangle, is $\geqslant (n-k)^n n^{-n} n!$ by van der Waerden's conjecture. By Minc's conjecture, we also have $\operatorname{per}(B) \leqslant ((n-k)!)^{n/(n-k)}$. If we let $L(k,n)$ denote the number of $k \times n$ Latin rectangles, then $L(1,n) = n!$, and the bounds derived above for the number of ways to extend any given rectangle give

$$L(k,n) \geqslant \prod_{j=0}^{k-1} \{(n-j)^n n^{-n} n!\} = n^{-kn}(n!)^{2n}((n-k)!)^{-n} , \tag{16.1}$$

$$L(k,n) \leqslant \prod_{j=0}^{k-1} \{(n-j)!\}^{n/(n-j)} . \tag{16.2}$$

Sharper estimates for $L(k,n)$ have been obtained through more powerful and complicated methods by Godsil and McKay (1990). They obtain an asymptotic relation for $L(k,n)$ that is valid for $k = \mathrm{o}(n^{6/7})$, and improved estimates for other k. [It is known that for any fixed k, the sequence $L(k,n)$ satisfies a linear recurrence with polynomial coefficients (Gessel 1987).] ⊠

There are problems in which inequalities for permanents give the correct asymptotic estimates. One such example is presented in Penrice (1991) which discusses a variation on the "problème des rencontres".

16.2. Probability theory and branching process methods

Many combinatorial enumeration results can be phrased in probabilistic language, and a few probabilistic techniques have appeared in the preceding sections. However, the stress throughout this chapter has been on elementary and generating-function approaches to asymptotic enumeration problems. Probabilistic methods provide another way to approach many of these problems. This has been appreciated more in the former Soviet Union than in the West, as can be seen in the books Kolchin (1986), Kolchin et al. (1978), and Sachkov (1978).

The last few years have seen a great increase in the applications of probabilistic methods to combinatorial enumeration and analysis of algorithms. Many powerful tools, such as martingales, branching processes, and Brownian motion asymptotics have been brought to bear on this topic. General introductions and references to these topics can be found in chapter 33 as well as in Aldous (1989), Alon and Spencer (1992), Arratia and Tavaré (1992b, 1994), Barbour et al. (1992), Devroye (1986, 1987), Erdős and Spencer (1974), Louchard (1983, 1986), Louchard et al. (1992), and Mahmoud (1992).

16.3. Statistical physics

There is an extensive literature in mathematical physics concerned with asymptotic enumeration, especially in Ising models of statistical mechanics and percolation methods. Many of the methods are related to combinatorial enumeration. For an introduction to them, see chapter 37 or the books Baxter (1982) and Kesten (1982).

16.4. Classical applied mathematics

There are many techniques, such as the ray method and the WKB method, that have been developed for solving differential and integral equations in what we might call classical applied mathematics. An introduction to them can be found in Bender and Orszag (1978). They are powerful, but they have the disadvantage that most of them are not rigorous, since they make assumptions about the form or the stability of the solution that are likely to be true, but have not been established. Therefore we have not presented such methods in this survey. For some examples of the nonrigorous applications of these methods to asymptotic enumeration, see the papers of Knessl and Keller (1990, 1991). It is likely that with additional work, more of these methods will be made rigorous, which will increase their utility.

17. Algorithmic and automated asymptotics

Deriving asymptotic expansions often involves a substantial amount of tedious work. However, much of it can now be done by computer symbolic algebra systems such as Macsyma, Maple, and Mathematica. There are many widely available packages that can compute Taylor-series expansions. Several can also compute certain types of limits, and some have implemented Gosper's indefinite hypergeometric summation algorithm (Gosper 1978). They ease the burden of carrying out the necessary but uninteresting parts of asymptotic analysis. They are especially useful in the exploratory part of research, when looking for identities, formulating conjectures, or searching for counterexamples.

Much more powerful systems are being developed. Given a sequence, there are algorithms that attempt to guess the generating function of that sequence

(Bergeron and Plouffe 1992, Getu et al. 1992). It is possible to go much further than that. Many of the asymptotic results in this chapter are stated in explicit forms. As an example, the asymptotics of a linear recurrence is derived easily from the characteristic polynomial and the initial conditions, as was shown in section 9.1. One needs to compute the roots of the characteristic polynomial, and that is precisely what computer systems do well. It is therefore possible to write programs that will derive the asymptotics behavior from the specification of the recurrence. More generally, one can analyze asymptotics of a much greater variety of generating functions. Flajolet, Salvy, and Zimmermann (Flajolet 1992, Flajolet et al. 1991b) have written a powerful program for just such computations. Their system uses Maple to carry out most of the basic analytic computations. It contains a remarkable amount of automated expertise in recognizing generating functions, computing their singularities, and extracting asymptotic information about their coefficients. For example, if

$$f(z) = -\log[1 + z\log(1 - z^2)] + (1 - z^3)^{-5} + \exp(z\,e^z) \,, \tag{17.1}$$

then the Flajolet–Salvy–Zimmermann system can determine that the singularity of $f(z)$ that is closest to the origin is at $z = \rho$, where ρ is the smallest positive root of

$$1 = -\rho\log(1 - \rho^2) \,, \tag{17.2}$$

and then can deduce that

$$[z^n]f(z) = n^{-1}\rho^{-n} + O(n^{-2}\rho^{-n}) \quad \text{as } n \to \infty \,. \tag{17.3}$$

The Flajolet–Salvy–Zimmermann system is even more powerful than indicated above, since it does not always require an explicit presentation of the generating function. Instead, often it can accept a formal description of an algorithm or data structure, derive the generating function from that, and then obtain the desired asymptotic information. For example, it can show that the average path length in a general planar tree with n nodes is

$$\frac{1}{2}\pi^{1/2}n^{3/2} + \frac{1}{2}n + O(n^{1/2}) \quad \text{as } n \to \infty \,. \tag{17.4}$$

What makes systems such as that of Flajolet et al. (1991b) possible is the phenomenon, already mentioned in section 6, that many common combinatorial operations on sets, such as unions and permutations, correspond in natural ways to operations on generating functions.

Further work extending that of Flajolet et al. (1991b) is undoubtedly going to be carried out. There are some basic limitations coming from the undecidability of even simple problems of arithmetic, which are already known to impose a limitation on the theories of indefinite integration. If we approximate a sum by an integral

$$\int_a^b x^{-\alpha}\,dx \,, \tag{17.5}$$

then as a next step we need to decide whether $\alpha = 1$ or not, since if $\alpha = 1$, this integral is $\log(b/a)$ (assuming $0 < a < b < \infty$), whereas if $\alpha \neq 1$, it is $(b^{1-\alpha} - a^{1-\alpha})/(1 - \alpha)$. Deciding whether $\alpha = 1$ or not, when α is given implicitly or by complicated expressions, can be arbitrarily complicated. However, such difficulties are infrequent, and so one can expect substantial increase in the applicability of automated systems for asymptotic analysis.

The question of decidability of asymptotic problems and generic properties of combinatorial structures that can be specified in various logical frameworks has been treated by Compton (1987, 1988, 1989). There is the beautiful recent theory of 0–1 laws for random graphs, which says that certain (so-called first-order) properties are true with probability either 0 or 1 for random graphs. Compton proves that certain classes of asymptotic theories also have 0–1 laws, and describes general properties that have to hold for almost all random structures in certain classes. His analysis uses Tauberian theorems and Hayman admissibility to determine asymptotic behavior. For some further developments in this area, see also Bender et al. (1992).

18. Guide to the literature

This section presents additional sources of information on asymptotic methods in enumeration and analysis of algorithms. It is not meant to be exhaustive, but is intended to be used as a guide in searching for methods and results. Many references have been presented already throughout this chapter. Here we describe only books that cover large areas relevant to our subject.

An excellent introduction to the basic asymptotic techniques is given in Graham et al. (1989). That book, intended to be an undergraduate textbook, is much more detailed than this chapter, and assumes no knowledge of asymptotics, but covers fewer methods. A less comprehensive and less elementary book that is oriented towards analysis of algorithms, but provides a good introduction to many asymptotic enumeration methods, is Greene and Knuth (1982).

The best source from which to learn the basics of more advanced methods, including many of those covered in this chapter, is de Bruijn (1958). It was not intended particularly for those interested in asymptotic enumeration, but almost all the methods in it are relevant. De Bruijn's volume is extremely clear, and provides insight into why and how various methods work.

General presentations of asymptotic methods, although usually with emphasis on applications to applied mathematics (differential equations, special functions, and so on) are available in the books Bleistein and Handelsman (1975), Erdélyi (1956), Fedoryuk (1987, 1989a), Olver (1974), Sirovich (1971), Szegö (1959), Wasow (1965), Wimp (1984), and Wong (1989). Integral transforms are treated extensively in Davies (1978), Doetsch (1955), Fedoryuk (1989b), Oberhettinger (1974), and Titchmarsh (1948). Books that deal with asymptotics arising in the analysis of algorithms or probabilistic methods include Alon and Spencer (1992), Bollobás

(1985), Erdős and Spencer (1974), Hofri (1987), Kemp (1984), Kolchin (1986), Kolchin et al. (1978), Mahmoud (1992), and Sachkov (1978).

Nice general introductions to combinatorial identities, generating functions, and related topics are presented in Comtet (1974), Stanley (1986), and Wilf (1990). Further material can be found in Andrews (1976), David and Barton (1962), Egorychev (1984), Goulden and Jackson (1983), Harary and Palmer (1973), Riordan (1958, 1968).

A very useful book is the compilation, Gonnet and Baeza–Yates (1991). While it does not discuss methods in too much detail, it lists a wide variety of enumerative results on algorithms and data structures, and gives references where the proofs can be found.

Last, but not least in our listing, is the three-volume work (Knuth 1973a,b, 1981). While it is devoted primarily to analysis of algorithms, it contains an enormous amount of material on combinatorics, especially asymptotic enumeration.

Acknowledgements

The author thanks R. Arratia, E. A. Bender, E. R. Canfield, H. Cohen, P. Flajolet, Z. Gao, D. E. Knuth, V. Privman, L. B. Richmond, E. Schmutz, N. J. A. Sloane, D. Stark, S. Tavaré, H. S. Wilf, D. Zagier and D. Zeilberger for their helpful comments on preliminary drafts of this chapter.

References

Abramowitz, M., and I.A. Stegun

[1970] eds., *Handbook of Mathematical Functions*, 9th printing (National Bureau of Standards/US Govt. Printing Office, Washington, DC).

Aczél, J.

[1966] *Lectures on Functional Equations and Their Applications* (Academic Press, New York).

Adams, C.R.

[1928] On the irregular cases of linear ordinary difference equations, *Trans. Amer. Math. Soc.* **30**, 507–541.

Aho, A.V., and N.J.A. Sloane

[1973] Some doubly exponential sequences, *Fibonacci Quart.* **11**, 429–437.

Aizenberg, J.A., and A.P. Yuzhakov

[1983] *Integral Representations and Residues in Multidimensional Complex Analysis, Translations of Mathematical Monographs*, Vol. 58, (American Mathematical Society, Providence, RI).

Aldous, D.

[1989] *Probability Approximations via the Poisson Clumping Heuristic* (Springer, Berlin).

Allouche, J.-P., and J. Shallit

[1992] The ring of k-regular sequences, *Theor. Comput. Sci.* **98**, 163–197.

Almkvist, G.

[1989] Proof of a conjecture about unimodal polynomials, *J. Number Theory* **32**, 43–57.

[1991] Exact asymptotic formulas for the coefficients of nonmodular functions, *J. Number Theory* **38**, 145–160.

[1993] A rather exact formula for the number of plane partitions, in: *A Tribute to Emil Grosswald: Number Theory and Related Analysis*, eds. M. Knopp and M. Sheingorn, *Contemp. Math.* **143**, 21–26.

Almkvist, G., and G.E. Andrews
[1991] A Hardy–Ramanujan–Rademacher formula for restricted partitions, *J. Number Theory* **38**, 135–144.

Alon, N., and J.H. Spencer
[1992] *The Probabilistic Method* (Wiley, New York).

Andrews, G.E.
[1974] Applications of basic hypergeometric functions, *SIAM Rev.* **16**, 441–484.
[1976] *The Theory of Partitions* (Addison-Wesley, Reading, MA).

Apostol, T.M.
[1957] *Mathematical Analysis* (Addison-Wesley, Reading, MA).
[1976] *Introduction to Analytic Number Theory* (Springer, Berlin).

Arney, J., and E.D. Bender
[1982] Random mappings with constraints on coalescence and number of origins, *Pacific J. Math.* **103**, 269–294.

Arratia, R., and S. Tavaré
[1992a] The cycle structure of random permutations, *Ann. Probab.* **20**, 1567–1591.
[1992b] Limit theorems for combinatorial structures via discrete process approximation, *Random Structures Algebra* **3**, 321–345.
[1994] Independent process approximations for random combinatorial structures, *Adv. in Math.* **104**, 90–154.

Arratia, R., L. Goldstein and L. Gordon
[1990a] Poisson approximation and the Chen–Stein method, *Statist. Sci.* **5**, 402–423.

Arratia, R., L. Gordon and M.S. Waterman
[1990b] The Erdős–Rényi law in distribution for coin tossing and sequence matching, *Ann. Statist.* **18**, 539–570.

Auluck, F.C., and C.B. Haselgrove
[1952] On Ingham's Tauberian theorem for partitions, *Proc. Cambridge Philos. Soc.* **48**, 566–570.

Ayoub, R.
[1963] *An Introduction to the Analytic Theory of Numbers* (American Mathematical Society, Providence, RI).

Baeza-Yates, R., R. Casas, J. Diaz and C. Martinez
[1992] On the average size of the intersection of binary trees, *SIAM J. Comput.* **21**, 24–32.

Baeza-Yates, R.A.
[1989] A trivial algorithm whose analysis is not: a continuation, *BIT* **29**, 378–394.

Bang, T.
[1976] Om matrixfunktioner som med et numerisk lille deficit viser v.d. Waerdens permanenthypotese, *Proc. 1976 Turku Scand. Math. Congress.*

Barbour, A.D., L. Holst and S. Janson
[1992] *Poisson Approximation* (Oxford University Press, Oxford).

Barnes, E.W.
[1904] On the homogeneous linear difference equation of the second order with linear coefficients, *Messenger Math.* **34**, 52–71.

Batchelder, P.M.
[1927] *An Introduction to Linear Difference Equations* (Harvard University Press). Reprinted: 1967 (Dover, New York).

Baxter, R.J.
[1982] *Exactly Solved Models in Statistical Mechanics* (Academic Press, New York).

Bender, C.M., and S.A. Orszag
[1987] *Applied Mathematical Methods for Scientists and Engineers* (McGraw-Hill, New York).

Bender, E.A.
[1973] Central and local limit theorem applied to asymptotic enumeration, *J. Combin. Theory A* **15**, 91–111.
[1974] Asymptotic methods in enumeration, *SIAM Rev.* **16**, 485–515.
[1975] An asymptotic expansion for the coefficients of some formal power series, *J. London Math. Soc.* 9, 451–458.

Bender, E.A., and J.R. Goldman

[1971] Enumerative uses of generating functions, *Indiana Univ. Math. J.* **20**, 753–765.

Bender, E.A., and L.B. Richmond

[1983] Central and local limit theorems applied to asymptotic enumeration. II: Multivariate generating functions, *J. Combin. Theory B* **34**, 255–265.

[1984] An asymptotic expansion for the coefficients of some power series II: Lagrange inversion, *Discrete Math.* **50**, 135–142.

[1986] A survey of the asymptotic behaviour of maps, *J. Combin Theory B* **40**, 297–329.

Bender, E.A., and S.G. Williamson

[1991] *Foundations of Applied Combinatorics* (Addison-Wesley, Reading, MA).

Bender, E.A., L.B. Richmond and S.G. Williamson

[1983] Central and local limit theorems applied to asymptotic enumeration. III. Matrix recursions, *J. Combin. Theory A* **35**, 263–278.

Bender, E.A., A.M. Odlyzko and L.B. Richmond

[1985] The asymptotic number of irreducible partitions, *European J. Combin.* **6**, 1–6.

Bender, E.A., Z.-C. Gao and L.B. Richmond

[1992] Submaps of maps. I. General 0–1 laws, *J. Combin. Theory B* **55**, 104–117.

Bergeron, F., and G. Cartier

[1988] Darwin: Computer algebra and enumerative combinatorics, in: *STACS-88, Lecture Notes in Computer Science,* Vol. 294, eds. R. Cori and M. Wirsing (Springer, Berlin) pp. 393–394.

Bergeron, F., and S. Plouffe

[1992] Computing the generating function of a series given its first few terms, *Exp. Math.* **1**, 307–312.

Bergeron, F., G. Labelle and P. Leroux

[1988] Functional equations for data structures, in: *STACS-88, Lecture Notes in Computer Science,* Vol. 294, eds. R. Cori and M. Wirsing (Springer, Berlin) pp. 73–80.

Berndt, B.C., and L. Schoenfeld

[1975/76] Periodic analogues of the Euler–Maclaurin and Poisson summation formulas with applications to number theory, *Acta Arithm.* **28**, 23–68.

Berry, M.V., and C.J. Howls

[1990] Hyperasymptotics, *Proc. R. Soc. London A* **430**, 653–667.

Bertozzi, A., and J. McKenna

[1993] Multidimensional residues, generating functions, and their application to queueing networks, *SIAM Rev.* **35**, 239–268.

Billingsley, P.

[1979] *Probability and Measure* (Wiley, New York).

Birkhoff, G.D.

[1911] General theory of linear difference equations, *Trans. Amer. Math. Soc.* **12**, 243–284.

[1930] Formal theory of irregular linear difference equations, *Acta Math.* **54**, 205–246.

Birkhoff, G.D., and W.J. Trjitzinsky

[1932] Analytic theory of singular difference equations, *Acta Math.* **60**, 1–89.

Bleistein, N., and R.A. Handelsman

[1975] *Asymptotic Expansions of Integrals,* 2nd Ed. (Holt, Rinehart & Winston, New York).

Bollobás, B.

[1985] *Random Graphs* (Academic Press, New York).

Brenti, F.

[1989] *Unimodal, Log-concave, and Pólya Frequency Sequences in Combinatorics, Mem. Amer. Math. Soc.* no. 413.

Brenti, F., G.F. Royle and D.G. Wagner

[1994] Location of zeros of chromatic and related polynomials of graphs, *Canad. J. Math.* **46**, 55–80.

Brigham, N.A.

[1950] A general asymptotic formula for partition functions, *Proc. Amer. Math. Soc.* **1**, 182–191.

Bromwich, T.I'a.

[1955] *An Introduction to the Theory of Infinite Series,* 2nd rev. Ed. (Macmillan, London).

Brown, G.G., and B.O. Shubert

[1984] On random binary trees, *Math. Oper. Res.* **9**, 43–65.

Canfield, E.R.

[1977] Central and local limit theorems for the coefficients of polynomials of binomial type, *J. Combin. Theory A* **23**, 275–290.

[1982] The asymptotic behavior of the Dickman–de Bruijn function, *Congress. Numerantium* **35**, 139–148.

[1984] Remarks on an asymptotic method in combinatorics, *J. Combin. Theory A* **37**, 348–352.

Car, M.

[1982] Factorisation dans $F_q[X]$, *C.R. Acad. Sci. Paris Série I* **294**, 147–150.

[1984] Ensembles de polynômes irréductibles et théorèmes de densité, *Acta Arithm.* **44**, 323–342.

Carlitz, L.

[1975] Permutations, sequences and special functions, *SIAM Rev.* **17**, 298–322.

Casas, R., D. Diaz and C. Martinez

[1991] Statistics on random trees, in: *Automata, Languages, and Programming, Proc. 18th ICALP, Madrid, 1991, Lecture Notes in Computer Science*, Vol. 510, eds. J. Leach Albert, B. Monien and M. Rodriguez Artalejo (Springer, Berlin) pp. 186–203.

Cassels, J.W.S.

[1960] On the representation of integers as the sums of distinct summands taken from a fixed set, *Acta Sci. Math. Hungar.* **21**, 111–124.

Cerlienco, L., M. Mignotte and F. Piras

[1987] Suites récurrentes linéaires, *L'Enseignement Math.* **33**, 67–108.

Charalambides, Ch.A., and A. Kyriakoussis

[1985] An asymptotic formula for the exponential polynomials and a central limit theorem for their coefficients, *Discrete Math.* **54**, 259–270.

Chen, L.H.Y.

[1975] Poisson approximation for dependent trials, *Ann. Probab.* **3**, 534–545.

Chung, F.R.K., R.L. Graham, J.A. Morrison and A.M. Odlyzko

[1995] Pebbling a chessboard, *Amer. Math. Monthly* **102**, 113–123.

Compton, K.J.

[1987] A logical approach to asymptotic combinatorics. I. First order properties, *Adv. in Math.* **65**, 65–96.

[1988] 0–1 laws in logic and combinatorics, in: *Proc. NATO Advanced Study Institute on Algorithms and Order*, ed. I. Rival (Reidel, Dordrecht) pp. 353–383.

[1989] A logical approach to asymptotic combinatorics. II. Monadic second-order properties, *J. Combin. Theory A* **50**, 110–131.

Comtet, L.

[1968] Birecouvrements et birevêtements d'un ensemble fini, *Studia Sci. Math. Hungar.* **38**, 137–152.

[1974] *Advanced Combinatorics* (Reidel, Dordrecht).

Cooper, C.N., and R.E. Kennedy

[1988] A partial asymptotic formul for the Niven numbers, *Fibonacci Quart.* **26**, 163–168.

Coquet, J.

[1983] A summation formula related to binary digits, *Invent. Math.* **73**, 107–115.

Courant, R., and D. Hilbert

[1953/1962] *Methods of Mathematical Physics*, Vols. 1 (1953) and 2 (1962) (Interscience, New York).

Cusick, T.W.

[1989] Recurrences for sums of powers of binomial coefficients, *J. Combin. Theory A* **52**, 77–83.

Daniels, H.E.

[1954] Saddlepoint approximations in statistics, *Ann. Math. Statist.* **25**, 631–650.

Darboux, G.

[1878] Mémoire sur l'approximation des fonctions de très-grands nombres, et sur une classe étendue de développements en série, *J. Math. Pures Appl.* **4**, 5–56, 377–416.

David, F.N., and D.E. Barton

[1962] *Combinatorial Chance* (Griffin, London).

Davies, B.
[1978] *Integral Transforms and Their Applications* (Springer, Berlin).

de Bruijn, N.G.
[1948] On Mahler's partition problem, *Indag. Math.* **10**, 210–220.
[1953] The difference-differential equation $F'(x) = e^{\alpha x+\beta} F(x-1)$, *Indag. Math.* **15**, 449–458.
[1958] *Asymptotic Methods in Analysis* (North-Holland, Amsterdam).

de Bruijn, N.G., D.E. Knuth and S.O. Rice
[1972] The average height of planted plane trees, in: *Graph Theory and Computing*, ed. R.-C. Read (Academic Press, New York) pp. 15–22.

Denef, J., and L. Lipshitz
[1987] Algebraic power series and diagonals, *J. Number Theory* **26**, 46–67.

Devaney, R.L.
[1989] *An Introduction to Chaotic Dynamical Systems*, 2nd Ed. (Addison-Wesley, Reading. MA).

Devroye, L.
[1986] A note on the expected height of binary search trees, *J. ACM* **33**, 489–498.
[1987] Branching processes in the analysis of the heights of trees, *Acta Inform.* **24**, 277–298.

Diaconis, P., and D. Freedman
[1980] Finite exchangeable sequences, *Ann. Probab.* **8**, 745–764.

Doetsch, G.
[1955] *Handbuch der Laplace Transformation* (Birkhäuser, Basel).

Drmota, M.
[1994] Asymptotic distributions and a multivariate Darboux method in enumeration problems, *J. Combin. Theory A* **67**, 169–184.

Durrett, R.
[1991] *Probability: Theory and Examples* (Wadsworth and Brooks/Cole, Pacific Grove, CA).

Egorychev, G.P.
[1981] The solution of van der Waerden's problem for permanents, *Adv. in Math.* **42**, 299–305.
[1984] *Integral Representation and the Computation of Combinatorial Sums* (American Mathematical Society, Providence, RI).

Erdélyi, A.
[1956] *Asymptotic Expansions* (Dover reprint).
[1961] General asymptotic expansions of Laplace integrals, *Arch. Rational Mech. Anal.* **7**, 1–20.

Erdélyi, A., and M. Wyman
[1963] The asymptotic evaluation of certain integrals, *Arch. Rational Mech. Anal.* **14**, 217–260.

Erdős, P.
[1941] On some asymptotic formulas in the theory of 'Factorisatio numerorum', *Ann. of Math.* **42**, 989–993. Corrections: 1943, 44, 647–651.

Erdős, P., and J. Lehner
[1941] The distribution of the number of summands in the partitions of a positive integer, *Duke Math. J.* **8**, 335–345.

Erdős, P., and J.H. Loxton
[1979] Some problems in partitio numerorum, *J. Austral. Math. Soc. A* **27**, 319–331.

Erdős, P., and B. Richmond
[1976] Concerning periodicity in the asymptotic behavior of partition functions, *J. Austral. Math. Soc. A* **21**, 447–456.

Erdős, P., and J. Spencer
[1974] *Probabilistic Methods in Combinatorics* (Academic Press/Akadémiai Kiadó, New York).

Erdős, P., and P. Turán
[1965] On some problems of a statistical group-theory, I, *Z. Wahrscheinlichkeitstheorie u. Verw. Gebiete* **4**, 175–186.
[1967a] On some problems of a statistical group-theory, II: *Acta Math. Acad. Sci. Hungar.* **18**, 151–163.
[1967b] On some problems of a statistical group-theory, III, *Acta Math. Acad. Sci. Hungar.* **18**, 309–320.

Erdős, P., and P. Turan

[1968] On some problems of a statistical group-theory, IV.: *Acta Math. Acad. Sci. Hungar.* **19**, 413–435.

Erdős, P., A. Hildebrand, A. Odlyzko, P. Pudaite and B. Reznick

[1987] The asymptotic behavior of a family of sequences, *Pacific J. Math.* **126**, 227–241.

Evgrafov, M.A.

[1961] *Asymptotic Estimates and Entire Functions* (Gordon and Breach, New York).

[1966] *Analytic Functions* (Dover, New York).

[1989] Series and integral representations, in: *Analysis I*, ed. R.V. Gamkrelidze (Springer, Berlin) pp. 1–81.

Falikman, D.I.

[1981] Proof of the van der Waerden conjecture on the permanent of a doubly stochastic matrix, *Mat. Zametki* **29**, 931–938 (in Russian).

Fedoryuk, M.V.

[1987] *Asymptotics: Integrals and Series* (Nauka, Moscow) (in Russian).

[1989a] Asymptotic methods in analysis, in: *Analysis I*, ed. R.V. Gamkrelidze (Springer, Berlin) pp. 83–191.

[1989b] Integral transforms, in: *Analysis I*, ed. R.V. Gamkrelidze (Springer, Berlin) pp. 193–232.

Feller, W.

[1968] *An Introduction to Probability Theory*, Vol. I, 3rd Ed. (Wiley, New York).

[1971] *An Introduction to Probability Theory*, Vol. II, 2nd Ed. (Wiley, New York).

Fields, J.L.

[1968] A uniform treatment of Darboux's method, *Arch. Rational Mech. Anal.* **27**, 289–305.

Fishburn, P.C., and A.M. Odlyzko

[1989] Unique subjective probability on finite sets, *J. Ramanujan Math. Soc.* 4, 1–23.

Fishburn, P.C., A.M. Odlyzko and F.S. Roberts

[1989] Two-sided generalized Fibonacci sequences, *Fibonacci Quart.* **27**, 352–361.

Flajolet, P.

[1978] *Analyse d'algorithmes de manipulation de fichiers*, publ. 321 (Institut de Recherche en Informatique et en Automatique, Rocquencourt).

[1980] Combinatorial aspects of continued fractions, *Discrete Math.* **32**, 125–161.

[1988] Mathematical methods in the analysis of algorithms and data structures, in: *Trends in Theoretical Computer Science*, ed. E. Börger (Computer Science Press) pp. 225–304.

[1992] Analytic analysis of algorithms, in: *Proc. ICALP '92, Springer Lecture Notes in Computer Science*, Vol. 623 (Springer, Berlin) pp. 186–210.

Flajolet, P., and J. Françon

[1989] Elliptic functions, continued fractions and doubled permutations, *European J. Combin.* **10**, 235–241.

Flajolet, P., and T. Lafforgue

[1994] Search costs in quadtrees and singularity perturbation asymptotics, *Discrete Comput. Geom.* **12**, 151–175.

Flajolet, P., and A.M. Odlyzko

[1982] The average height of binary trees and other simple trees, *J. Comput. System Sci.* **25**, 171–213.

[1984] Limit distributions for coefficients of iterates of polynomials with application to combinatorial enumeration, *Math. Proc. Cambridge Philos. Soc.* **96**, 237–253.

[1990a] Random mapping statistics, in: *Advances in Cryptology: Proc. Eurocrypt '89, Springer Lecture Notes in Computer Science*, Vol. 434, ed. J.-J. Quisquater (Springer, Berlin) pp. 329–354.

[1990b] Singularity analysis of generating function, *SIAM J. Discrete Math.* 3, 216–240.

Flajolet, P., and B. Richmond

[1992] Generalized digital trees and their difference-differential equations, *Random Structures Algor.* **3**, 305–320.

Flajolet, P., and R. Schott

[1990] Non-overlapping partitions, continued fractions, Bessel functions and a divergent series, *European J. Combin.* **11**, 421–432.

Flajolet, P., and R. Sedgewick

[1986] Digital search trees revisited, *SIAM J. Comput.* **15**, 748–767.

Flajolet, P., and M. Soria

[1990] Gaussian limiting distributions for the number of components in combinatorial structures, *J. Combin. Theory A* **53**, 165–182.

[1993] General combinatorial schemes with Gaussian limit distributions and exponential tails, *Discrete Math.* **114**, 159–180.

Flajolet, P., J.-C. Raoult and J. Vuillemin

[1979] The number of registers required to evaluate arithmetic expressions, *Theor. Comput. Sci.* **9**, 99–125.

Flajolet, P., M. Régnier and R. Sedgewick

[1985] Some uses of the Mellin integral transform in the analysis of algorithms, in: *Combinatorial Algorithms on Words*, eds. A. Apostolico and Z. Galil (Springer, Berlin) pp. 241–254.

Flajolet, P., P. Kirschenhofer and R. Tichy

[1988] Deviations from normality in random strings, *Probab. Theory Related Fields* **80**, 139–150.

Flajolet, P., G. Gonnet, C. Puech and J.M. Robson

[1991a] The analysis of multidimensional searching in quad-trees, in: *Proc. 2nd ACM-SIAM Symp. on Discrete Algorithms* (SIAM, Philadelphia, PA) pp. 100–109.

Flajolet, P., B. Salvy and P. Zimmermann

[1991b] Automatic average-case analysis of algorithms, *Theor. Comput. Sci.* **79**, 37–109.

Flajolet, P., Z. Gao, A.M. Odlyzko and B. Richmond

[1993a] The height of binary trees and other simple trees, *Combin. Probab. Comput.* **2**, 145–156.

Flajolet, P., G. Gonnet, C. Puech and J.M. Robson

[1993b] Analytic variations on quadtrees, *Algorithmica* **10**, 473–500.

Flajolet, P., P. Grabner, P. Kirschenhofer, H. Prodinger and R.F. Tichy

[1994] Mellin transforms and asymptotics: digital sums, *Theor. Comp. Sci.* **123**, 291–314.

Flatto, L.

[1989] The longer queue model, *Probab. Eng. Inform. Sci.* **3**, 537–559.

Flatto, L., and S. Hahn

[1984] Two parallel queues created by arrivals with two demands. I, *SIAM J. Appl. Math.* **44**, 1041–1053.

Ford, G.W., and G.E. Uhlenbeck

[1956a] Combinatorial problems in the theory of graphs I, *Proc. Nat. Acad. Sci. U.S.A.* **42**, 122–128.

[1956b] Combinatorial problems in the theory of graphs III, *Proc. Nat. Acad. Sci. U.S.A.* **42**, 529–535.

[1957] Combinatorial problems in the theory of graphs IV, *Proc. Nat. Acad. Sci. U.S.A.* **43**, 163–167.

Ford, G.W., G.E. Uhlenbeck and R.Z. Norman

[1956] Combinatorial problems in the theory of graphs II, *Proc. Nat. Acad. Sci. U.S.A.* **42**, 203–208.

Fredman, M.L., and D.E. Knuth

[1974] Recurrence relations based on minimization, *J. Math. Anal. Appl.* **48**, 534–[illegible]

Friedland, S.

[1979] A lower bound for the permanent of a doubly stochastic matrix, *Ann. of Math. (2)* **110**, 167–176.

Frieze, A.

[1991] On the length of the longest monotone subsequence in a random permutation, *Ann. Appl. Probab.* **1**, 301–305.

Fristedt, B.

[1993] The structure of random partitions of large integers, *Trans. Amer. Math. Soc.* **337**, 703–735.

Furstenberg, H.

[1967] Algebraic function fields over finite fields, *J. Algebra* **7**, 271–272.

Galambos, J.

[1977] Bonferroni inequalities, *Ann. Probab.* **5**, 577–581.

Galambos, J., and Y. Xu

[1993] Some optimal bivariate Bonferroni-type bounds, *Proc. Amer. Math. Soc.* **117**, 523–528.

Ganelius, T.H.

[1971] *Tauberian Remainder Theorems, Lecture Notes in Mathematics*, Vol. 232 (Springer, Berlin).

Gao, Z., and L.B. Richmond

[1992] Central and local limit theorems applied to asymptotic enumeration. IV: Multivariate generating functions, *J. Appl. Comput. Anal.* **41**, 177–186.

Gardy, D.

[1992] Méthodes de col et lois limités en analyse combinatoire, *Theor. Comput. Sci.* 94, 261–280.

[1995] Some results on the asymptotic behavior of coefficients of large powers of functions, *Discrete Math.*, to be published.

Gardy, D., and P. Solé

[1992] Saddle point techniques in asymptotic coding theory, in: *Algebraic Coding, Proc. 1st French–Soviet Workshop, 1991, Lecture Notes in Computer Science,* Vol. 573, eds. G. Cohen, S. Litsyn, A. Lobstein and G. Zémor (Springer, Berlin) pp. 75–81.

Garsia, A.M., and S.A. Joni

[1977] A new expansion for umbral operators and power series inversion, *Proc. Amer. Math. Soc.* **64**, 179–185.

Gessel, I.M.

[1987] Counting Latin rectangles, *Bull. Amer. Math. Soc.* **16**, 79–82.

[1990] Symmetric functions and P-recursiveness, *J. Combin. Theory A* **53**, 257–286.

Getu, S., L.W. Shapiro, W.-J. Woan and L.C. Woodson

[1992] How to guess a generating function, *SIAM J. Discrete Math.* **5**, 497–499.

Godsil, C.D., and B.D. McKay

[1990] Asymptotic enumeration of Latin rectangles, *J. Combin. Theory B* **48**, 19–44.

Goh, W.M.Y., and E. Schmutz

[1991a] The expected order of a random permutation, *Bull. London Math. Soc.* **23**, 34–42.

[1991b] A central limit theorem on $\mathrm{GL}_n(F_q)$, *Rand. Struct. Algebra* **2**, 47–53.

[1995] Distribution of the number of distinct parts in a random partition, *J. Combin. Theory A*, to appear.

Goncharov, V.L.

[1944] From the domain of combinatorial analysis, *Izv. Akad. Nauk SSSR Ser. Math.* **8**(1), 3–48 [1962, *Transl. Amer. Math. Soc.* 19, 1–46].

Gonnet, G.H., and R. Baeza-Yates

[1991] *Handbook of Algorithms and Data Structures,* 2nd Ed. (Addison-Wesley, Reading, MA).

Good, I.J.

[1960] Generalizations to several variables of Lagrange's expansion, with applications to stochastic processes, *Proc. Cambridge Philos. Soc.* 56, 367–380.

Gordon, B., and L. Houten

[1969] Notes on plane partitions. III, *Duke Math. J.* **26**, 801–824.

Gosper Jr, R.W.

[1978] Decision procedure for indefinite hypergeometric summation, *Proc. Nat. Acad. Sci. U.S.A.* **75**, 40–42.

Gould, H.W.

[1972] *Combinatorial Identities* (private printing).

Goulden, I.P., and D.M. Jackson

[1983] *Combinatorial Enumeration* (Wiley, New York).

Gradshteyn, I.S., and I.M. Ryzhik

[1965] *Table of Integrals, Series and Products* (Academic Press, New York).

Graham, R.L., D.E. Knuth and O. Patashnik

[1989] *Concrete Mathematics* (Addison Wesley, Reading, MA).

Greenberg, A.G., B.D. Lubachevsky and A.M. Odlyzko

[1988] Simple, efficient asynchronous parallel algorithms for maximization, *ACM Trans. Programming Languages Systems,* pp. 313–337.

Greene, D.H., and D.E. Knuth

[1982] *Mathematics for the Analysis of Algorithms,* 2nd Ed. (Birkhäuser, Boston).

Griggs, J.R., P.J. Hanlon, A.M. Odlyzko and M.S. Waterman

[1990] On the number of alignments of k sequences, *Graphs Combin.* **6**, 133–146.

Grosswald, E.

[1966] Generalization of a formula of Hayman and its application to the study of Riemann's zeta function, *Illinois J. Math.* **10**, 9–23. Correction: 1969, **13**, 276–280.

Guibas, L.J., and A.M. Odlyzko

[1978] Maximal prefix-synchronized codes, *SIAM J. Appl. Math.* **35**, 401–418.

[1980] Long repetitive patterns in random sequences, *Z. Wahrscheinlichkeitstheorie u. Verw. Gebiete* **53**, 241–262.

[1981] String overlaps, pattern matching, and nontransitive games, *J. Combin. Theory A* **30**, 183–208.

Gutjahr, W.J.

[1992] The variance of level numbers in certain families of trees, *Random Structures Algebra* **3**, 361–374.

Halton, J.H.

[1989] The properties of random trees, *Inform. Sci.* **47**, 95–133.

Handelsman, R.A., and J.S. Lew

[1970] Asymptotic expansion of Laplace transforms near the origin, *SIAM J. Math. Analysis* **1**.

Hansen, E.R.

[1975] *A Table of Series and Products* (Prentice-Hall, Englewood Cliffs, NJ).

Hansen, J.

[1994] Order statistics for decomposable combinatorial structures, *Rand. Struct. Algebra* **5**, 517–533.

Harary, F., and E.M. Palmer

[1973] *Graphical Enumeration* (Academic Press, New York).

Harary, F., R.W. Robinson and A.J. Schwenk

[1975] Twenty-step algorithm for determining the asymptotic number of trees of various species, *J. Austral. Math. Soc. Ser. A* **20**, 483–503.

Hardy, G.H.

[1949] *Divergent Series* (Oxford University Press, London).

Hardy, G.H., and J.E. Littlewood

[1914a] Tauberian theorems concerning power series and Dirichlet's series whose coefficients are positive, *Proc. London Math. Soc.* (2) **13**, 174–191. Reprinted: 1974, *Collected Papers of G.H. Hardy,* Vol. 6, pp. 510–527.

[1914b] Some theorems concerning Dirichlet's series, *Messenger Math.* **43**, 134–147. Reprinted: 1974, *Collected Papers of G.H. Hardy,* Vol. 6, pp. 542–555.

Hardy, G.H., and S. Ramanujan

[1917] Asymptotic formulae for the distribution of integers of various types, *Proc. London Math. Soc.* (2)**16**, 112–132. Reprinted: 1966, *Collected Papers of G.H. Hardy,* Vol. 1, pp. 277–293.

Hardy, G.H., J.E. Littlewood and G. Pólya

[1952] *Inequalities*, 2nd Ed. (Cambridge University Press, Cambridge).

Harper, L.H.

[1967] Stirling behavior is asymptotically normal, *Ann. Math. Statist.* **38**, 410–414.

Harris, B.

[1960] Probability distributions related to random mappings, *Ann. Math. Statist.* **31**, 1042–1062.

Harris, B., and C.J. Park

[1971] The distribution of linear combinations of the sample occupancy numbers, *Nederl. Akad. Wetensch. Proc. Ser. A* **74** [= *Indag. Math.* **33**], 121–134.

Harris, B., and L. Schoenfeld

[1968] Asymptotic expansions for the coefficients of analytic functions, *Illinois J. Math.* **12**, 264–277.

Harris, T.E.

[1963] *The Theory of Branching Processes* (Springer, Berlin).

Harris, W.A., and Y. Sibuya

[1964] Asymptotic solutions of systems of nonlinear difference equations, *Arch. Rational Mech. Anal.* **15**, 277–395.

[1965] General solution of nonlinear difference equations, *Trans. Amer. Math. Soc.* **115**, 62–75.

Haselgrove, C.B., and H.N.V. Temperley

[1954] Asymptotic formulae in the theory of partitions, *Proc. Cambridge Philos. Soc.* **50**, 225–241.

Hautus, M.L.J., and D.A. Klarner

[1971] The diagonals of a double power series, *Duke Math. J.* **38**, 229–235.

Hayman, W.K.

[1956] A generalization of Stirling's formula, *J. Reine Angew. Math.* **196**, 67–95.

Henrici, P.

[1974–1986] *Applied and Computational Complex Analysis*, Vols. 1 (1974), 2 (1977), 3 (1986) (Wiley, New York).

Hille, E.

[1969] *Lectures on Ordinary Differential Equations* (Addison-Wesley, Reading, MA).

[1976] *Ordinary Differential Equations in the Complex Domain* (Wiley, New York).

Hofbauer, J.J.

[1979] A short proof of the Lagrange–Good formula, *Discrete Math.* **25**, 135–139.

Hofri, M.

[1987] *Probabilistic Analysis of Algorithms* (Springer, Berlin).

Hunter, C.

[1968] Asymptotic solutions of certain linear difference equations, with applications to some eigenvalue problems, *J. Math. Anal. Appl.* **24**, 279–289.

Immink, G.K.

[1984] *Asymptotics of Analytic Difference Equations, Lecture Notes in Mathematics*, Vol. 1085 (Springer, Berlin).

Ingham, A.E.

[1941] A Tauberian theorem for partitions, *Ann. of Math.* **42**, 1075–1090.

Jacquet, P., and M. Régnier

[1986] Trie partitioning process: limiting distributions, in: *CAAP '86, Lecture Notes in Computer Science*, Vol. 214, ed. P. Franchi-Zannettacci (Springer, Berlin) pp. 196–210.

[1988] Normal limiting distribution of the size of tries, in: *Performance '87*, eds. P.-J. Courtois and G. Latouche (North-Holland, Amsterdam) pp. 209–223.

[1995] Normal limiting distribution for the size and the external path length of tries, in preparation.

Jolley, L.B.W.

[1961] *Summation of Series*, 2nd Ed. (Dover, New York).

Jonassen, A.T., and D.E. Knuth

[1978] A trivial algorithm whose analysis is not, *J. Comput. Sys. Sci.* **16**, 301–322.

Jones, W.B., and W.J. Thron

[1980] *Continued Fractions: Analytic Theory and Applications* (Addison-Wesley, Reading, MA).

Jungen, R.

[1931] Sur les séries de Taylor n'ayant que des singularitiés algébrico-logarithmiques sur leur cercle de convergence, *Comment. Math. Helv.* **3**, 266–306.

Kapoor, S., and E.M. Reingold

[1985] Recurrence relations based on minimization and maximization, *J. Math. Anal. Appl.* **109**, 591–604.

Karlin, S.

[1968] *Total Positivity*, Vol. 1 (Stanford University Press, Stanford, CA).

Karp, R.M.

[1991] Probabilistic recurrence relations, in: *Proc. 23rd ACM Symp. on Theory of Computing* (ACM, New York) pp. 190–197.

Kemp, R.

[1984] *Fundamentals of the Average Case Analysis of Particular Algorithms* (Wiley, New York).

[1987] A note on the number of leftist trees, *Inform. Process. Lett.* **25**, 227–232.

[1990] Further results on leftist trees, in: *Random Graphs '87*, eds. M. Karónski, J. Jaworski and A. Rucinski (Wiley, New York) pp. 103–130.

Kesten, H.

[1982] *Percolation Theory for Mathematicians* (Birkhäuser, Basel).

Kirschenhofer, P.

[1987] A tree enumeration problem involving the asymptotics of the 'diagonals' of a power series, *Ann. Discrete Math.* **33**, 157–170.

Kirschenhofer, P., and H. Prodinger

[1991] On some applications of formulae of Ramanujan in the analysis of algorithms, *Mathematika* **38**, 14–33.

Klarner, D.A.

[1968] A combinatorial formula involving the Fredholm integral equation, *J. Combin. Theory* **5**, 59–74.

Klarner, D.A., and R.L. Rivest

[1974] Asymptotic bounds for the number of convex n-ominoes, *Discrete Math.* **8**, 31–40.

Knessl, C., and J.B. Keller

[1990] Partition asymptotics for recursion equations, *SIAM J. Appl. Math.* **50**, 323–338.

[1991] Stirling number asymptotics from recursion equations using the ray method, *Studia Appl. Math.* **84**, 43–56.

Knopfmacher, A., A.M. Odlyzko, B. Richmond, G. Szekeres and N. Wormald

[1995] Manuscript, in preparation.

Knopp, K.

[1971] *Theory and Application of Infinite Series*, 2nd Ed. Reprinted: Hafner.

Knuth, D.E.

[1973a] *The Art of Computer Programming*, Vol. 1: *Fundamental Algorithms*, 2nd Ed. (Addison-Wesley, Reading, MA).

[1973b] *The Art of Computer Programming*, Vol. 3: *Sorting and Searching* (Addison Wesley, Reading, MA).

[1981] *The Art of Computer Programming*, Vol. 2: *Semi-Numerical Algorithms*, 2nd Ed. (Addison-Wesley, Reading, MA).

Knuth, D.E., and B. Pittel

[1989] A recurrence related to trees, *Proc. Amer. Math. Soc.* **105**, 335–349.

Knuth, D.E., and A. Schönhage

[1978] The expected linearity of a simple equivalence algorithm, *Theor. Comput. Sci.* **6**, 281–315.

Kolchin, V.F.

[1986] *Random Mappings* (Optimization Software Inc., New York).

Kolchin, V.F., B.A. Sevast'yanov and V.P. Chistyakov

[1978] *Random Allocations* (Wiley, New York).

Komlós, J., A.M. Odlyzko, L.H. Ozarow and L.A. Shepp

[1990] On the properties of a tree-structured server process, *Ann. Appl. Probab.* **1**, 118–125.

Kooman, R.J.

[1989] *Convergence properties of recurrence sequences*, Ph.D. Dissertation (University of Leiden).

Kooman, R.J., and R. Tijdeman

[1990] Convergence properties of linear recurrence sequences, *Nieuw Arch. Wisk. (4)* **4**, 13–25.

Kruskal, M.D.

[1954] The expected number of components under a random mapping function, *Amer. Math. Monthly* **61**, 392–397.

Kuczma, M.

[1968] *Functional Equations in a Single Variable* (Polish Scientific Publishers, Warsaw).

Labelle, G.

[1981] Une nouvelle démonstration combinatoire des formules d'inversion de Lagrange, *Adv. in Math.* **42**, 217–247.

Lagarias, J.C., A.M. Odlyzko and D.B. Zagier

[1985] On the capacity of disjointly shared networks, *Computer Networks and ISDN Systems* **10**, 275–285.

Lee, M.-Y.

[1992] Bivariate Bonferroni inequalities, *Aequationes Math.* **44**, 220–225.

Leray, J.

[1959] Le calcul différentiel et intégral sur une variété analytique complexe, *Bull. Soc. Math. France* **87**, 81–180.

Lewin, L.

[1981] *Polylogarithms and Associated Functions* (North-Holland, Amsterdam).

Lichtin, B.

[1991] The asymptotics of a lattice point problem associated to a finite number of polynomials. I, *Duke Math. J.* **63**, 139–192.

Lipshitz, L.

[1988] The diagonal of a *D*-finite power series is *D*-finite, *J. Algebra* **113**, 373–378.

[1989] *D*-Finite power series, *J. Algebra* **122**, 353–373.

Lipshitz, L., and A. van der Poorten

[1990] Rational functions, diagonals, automata and arithmetic, in: *Number Theory*, ed. R.A. Mollin (Walter de Gruyter, Berlin) pp. 339–358.

Logan, B.F., and L.A. Shepp

[1977] A variational problem for random Young tableaux, *Adv. in Math.* **26**, 206–222.

Logan, B.F., J.E. Mazo, A.M. Odlyzko and L.A. Shepp

[1983] On the average product of Gauss–Markov variables, *Bell Sys. Tech. J.* **62**, 2993–3006.

Louchard, G.

[1983] The Brownian motion: a neglected tool for the complexity analysis of sorted table manipulation, *RAIRO Theor. Inform.* **17**, 365–385.

[1984] The Brownian excursion: a numerical analysis, *Comput. Math. Appl.* **10**, 413–417.

[1986] Brownian motion and algorithm complexity, *BIT* **26**, 17–34.

[1987] Exact and asymptotic distributions in digital and binary search trees, *RAIRO Inform. Théor. Appl.* **21**, 479–495.

Louchard, G., B. Randrianarimanana and R. Schott

[1992] Dynamic algorithms in D.E. Knuth's model; a probabilistic analysis, *Theor. Comput. Sci.* **93**, 201–255.

Łuczak, T.

[1994] The number of trees with a large diameter, to be published.

Lueker, G.S.

[1980] Some techniques for solving recurrences, *Comput. Surveys* **12**, 419–436.

Macintyre, A.J., and R. Wilson

[1954] Operational methods and the coefficients of certain power series, *Math. Ann.* **127**, 243–250.

Mahler, K.

[1940] On a special functional equation, *J. London Math. Soc.* **15**, 115–123.

[1976] On a class of nonlinear functional equations connected with modular functions, *J. Austral. Math. Soc. A* **22**, 65–118.

[1981] On a special nonlinear functional equation, *Proc. R. Soc. London A* **378**, 155–178.

[1983] On the analytic relation of certain functional and difference equations, *Proc. R. Soc. London A* **389**, 1–13.

Mahmoud, H.M., and B. Pittel

[1989] Analysis of the space of search trees under the random insertion algorithm, *J. Algorithms* **10**, 52–75.

Mahmoud, H.S.

[1992] *Evolution of Random Search Trees* (Wiley, New York).

Malgrange, B.

[1974] Sur les points singuliers des équations différentielles, *L'Enseignement Math.* **20**, 147–176.

Mallows, C.L., A.M. Odlyzko and N.J.A. Sloane

[1975] Upper bounds for modular form, lattices, and codes, *J. Algebra* **36**, 68–76.

Malyshev, V.A.

[1972] An analytic method in the theory of two-dimensional positive random walks, *Sibirsk. Mat. Zh.* **13**, 1314–1329 (in Russian).

Maté, A., and P. Nevai

[1984] Sublinear perturbations of the differential equation $y^{(n)} = 0$ and of the analogous difference equation, *J. Differential Equations* **53**, 234–257.

[1985] Asymptotics for solutions of smooth recurrence relations, *Proc. Amer. Math. Soc.* **93**, 423–429.

Mazo, J.E., and A.M. Odlyzko

[1990] Lattice points in high-dimensional spheres, *Monatsh. Math.* **110**, 47–61.

McKay, B.D.

[1990] The asymptotic numbers of regular tournaments, eulerian digraphs, and eulerian and oriented graphs, *Combinatorica* **10**, 367–377.

McKay, B.D., and N.C. Wormald

[1990] Asymptotic enumeration by degree sequence of graphs of high degree, *European J. Combin.* **11**, 565–580.

Meinardus, G.

[1954] Asymptotische Aussagen über Partitionen, *Math. Z.* **59**, 388–398.

Meir, A., and J.W. Moon

[1978] On the altitude of nodes in random trees, *Canadian J. Math.* **30**, 997–1015.

[1984] On random mapping patterns, *Combinatorica* **4**, 61–70.

[1987] Some asymptotic results useful in enumeration problems, *Aequationes Math.* **33**, 260–268.

[1989] On an asymptotic method in enumeration, *J. Combin. Theory A* **51**, 77–89.

[1990] The asymptotic behavior of coefficients of powers of certain generating functions, *European J. Combin.* **11**, 581–587.

Mendelsohn, N.S.

[1956] The asymptotic series for a certain class of permutation problems, *Canad. J. Math.* **8**, 234–244.

Milne-Thomson, L.M.

[1933] *The Calculus of Finite Differences* (MacMillan, New York).

Mitrinovic, D.S.

[1970] *Analytic Inequalities* (Springer, Berlin).

Moews, D.

[1995] Explicit Tauberian bounds for multivariate functions, to be published.

Moon, J.W.

[1970] *Counting Labeled Trees, Canad. Math. Monograph,* No. 1 (Canadian Mathematical Congress).

[1987] Some enumeration results on series–parallel networks, in: *Random Graphs '85,* eds. M. Karónski and Z. Palka, *Ann. Discrete Math.* 33, 199–226.

Moser, L., and M. Wyman

[1955] On the solutions of $x^d = 1$ in symmetric groups, *Canad. J. Math.* **7**, 159–168.

[1956] Asymptotic expansions, *Canad. J. Math.* **8**, 225–233.

[1957] Asymptotic expansions II, *Canad. J. Math.* **9**, 194–209.

[1958a] Stirling numbers of the second kind, *Duke Math. J.* **25**, 29–43.

[1958b] Asymptotic development of the Stirling numbers of the first kind, *J. London Math. Soc.* **33**, 133–146.

Nörlund, N.E.

[1924] *Vorlesungen über Differenzenrechnung* (Springer, Berlin). Reprinted: 1954 (Dover, New York).

Oberhettinger, F.

[1974] *Tables of Mellin Transforms* (Springer, Berlin).

Odlyzko, A.M.

[1982] Periodic oscillations of coefficients of power series that satisfy functional equations, *Adv. in Math.* **44**, 180–205.

[1984] Some new methods and results in tree enumeration, *Congress. Numerantium* **42**, 27–52.

[1985a] On heights of monotonically labelled binary trees, *Congress. Numerantium* **44**, 305–314.

[1985b] Enumeration of strings, in: *Combinatorial Algorithms on Words,* eds. A. Apostolico and Z. Galil (Springer, Berlin) pp. 205–228.

[1985c] Applications of symbolic mathematics to mathematics, in: *Applications of Computer Algebra,* ed. R. Pavalle (Kluwer, Dordrecht) pp. 95–111.

[1992] Explicit Tauberian estimates for functions with positive coefficients, *J. Comput. Appl. Math.* **41**, 187–197.

Odlyzko, A.M., and L.B. Richmond

[1980] On the compositions of an integer, in: *Combinatorial Mathematics VII, Lecture Notes in Mathematics,* Vol. 829, eds. R.-W. Robinson, G.W. Southern and W.D. Wallis (Springer, Berlin) pp. 119–210.

[1982] On the unimodality of some partition polynomials, *European J. Combin.* 3, 69–84.

[1985a] On the unimodality of high convolutions of discrete distributions, *Ann. Probab.* **13**, 299–306.

[1985b] On the number of distinct block sizes in partitions of a set, *J. Combin. Theory A* **38**, 170–181.

[1985c] Asymptotic expansions for the coefficients of analytic generating functions, *Aequationes Math.* **28**, 50–63.

Odlyzko, A.M., and H.S. Wilf

[1987] Bandwidths and profiles of trees, *J. Combin. Theory B* **42**, 348–370. Condensed summary of results: in: *Graph Theory and its Applications to Algorithms and Computer Science*, eds. Y. Alavi et al. (Wiley, 1985) pp. 605–622.

[1988] The editor's corner: *n* coins in a fountain, *Amer. Math. Monthly* **95**, 840–843.

[1991] Functional iteration and the Josephus problem, *Glasgow Math. J.* 33, 235–240.

Odlyzko, A.M., B. Poonen, H. Widom and H.S. Wilf

[1995] Manuscript, in preparation.

Olver, F.W.J.

[1974] *Asymptotics and Special Functions* (Academic Press, New York).

Otter, R.

[1948] The number of trees, *Ann. of Math.* 49, 583–599.

Pavlov, A.I.

[1992] On the number of substitutions with cycle lengths from a given set, *Discrete Appl. Math.* **2**, 445–459.

Penrice, S.G.

[1991] Derangements, permanents, and Christmas presents, *Amer. Math. Monthly* **98**, 617–620.

Perron, O.

[1957] *Die Lehre von den Kettenbruchen*, 3rd Ed. (Teubner, Stuttgart). Reprinted by Chelsea.

Pólya, G.

[1937] Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen, *Acta Math.* **68**, 145–254.

[1969] On the number of certain lattice polygons, *J. Combin. Theory* **6**, 102–105.

Pólya, G., and R.C. Read

[1987] *Combinatorial Enumeration of Groups, Graphs, and Chemical Compounds* (Springer, Berlin).

Pólya, G., and G. Szegö

[1972/1976] *Problems and Theorems in Analysis*, 2 volumes, English translation (Springer, Berlin).

Popken, A.

[1953] Asymptotic expansions from an algebraic standpoint, *Indag. Math.* **15**, 131–143.

Postnikov, A.G.

[1980] Tauberian theory and its applications, *Proc. Steklov Inst. Math.* **144** (English translation, *Amer. Math. Soc. Transl.*).

Privman, V., and N.M. Svrakic

[1988] Difference equations in statistical mechanics: I. Cluster statistics models; II. Solid-on-solid models in two dimensions, *J. Statist. Phys.* **51**, 1091–1110; 1111–1126.

[1989] *Directed Models of Polymers, Interfaces, and Clusters: Scaling and Finite-Size Properties, Lecture Notes in Physics*, Vol. 338 (Springer, Berlin).

Rademacher, H.

[1937] On the partition function, *Proc. London Math. Soc.* **13**, 241–254.

Regev, A.

[1981] Asymptotic values for degrees associated with strips of Young diagrams, *Adv. in Math.* **41**, 115–136.

Rényi, A.

[1962] Three more proofs and a generalization of a theorem of Irving Weiss, *Magyar Tud. Akad. Mat. Kutato Int. Közl.* **7**, 203–214.

Rényi, A., and G. Szekeres

[1967] On the height of trees, *J. Austral. Math. Soc.* 7, 497–507.

Richmond, L.B.

[1975a] Asymptotic relations for partitions, *J. Number Theory* 4, 389–405.

[1975b] The moments of partitions. II, *Acta Arithm.* **28**, 229–243.

[1976] Asymptotic relations for partitions, *Trans. Amer. Math. Soc.* **219**, 379–385.

Riordan, J.

[1958] *Introduction to Combinatorial Analysis* (Wiley, New York).

[1968] *Combinatorial Identities* (Wiley, New York).

Roth, K.F., and G. Szekeres

[1954] Some asymptotic formulae in the theory of partitions, *Quart. J. Math. Oxford Ser. 2* **5**, 241–259.

Sachkov, V.N.

[1978] *Probabilistic Methods in Combinatorial Analysis* (in Russian) (Nauka, Moscow).

Schmutz, E.

[1989] Asymptotic expansions for the coefficients of $e^{P(z)}$, *Bull. London Math. Soc.* **21**, 482–486.

Schrijver, A.

[1978] A short proof of Minc's conjecture, *J. Combin. Theory A* **25**, 80–83.

Sedgewick, R.

[1978] Data movement in odd–even merging, *SIAM J. Comput.* **7**, 239–272.

Shepp, L.A., and S.P. Lloyd

[1966] Ordered cycle lengths in a random permutation, *Trans. Amer. Math. Soc.* **121**, 340–357.

Shohat, J.A., and J.D. Tamarkin

[1943] *The Problem of Moments* (American Mathematical Society, Providence, RI).

Sirovich, L.

[1971] *Techniques of Asymptotic Analysis* (Springer, Berlin).

Sloane, N.J.A.

[1973] *A Handbook of Integer Sequences* (Academic Press, New York).

Sloane, N.J.A., and S. Plouffe

[1995] *The Encyclopedia of Integer Sequences* (Academic Press, New York) in press.

Sneddon, I.N.

[1972] *The Uses of Integral Transforms* (McGraw-Hill, New York).

Spencer, J.

[1987] *Ten Lectures on the Probabilistic Method* (SIAM, Philadelphia, PA).

Stanley, R.P.

[1978] Generating functions, in: *Studies in Combinatorics, M.A.A. Studies in Mathematics,* Vol. 17, ed. G.-C. Rota (Mathematical Association of America, New York) pp. 100–141.

[1980] Differentiably finite power series, *European J. Combin.* **1**, 175–188.

[1986] *Enumerative Combinatorics* (Wadsworth and Brooks/Cole, Monterey, CA).

[1989] Log-concave and unimodal sequences in algebra, combinatorics, and geometry, in: *Graph Theory and its Applications: East and West, Ann. New York Acad. Sci.* **576**, 500–535.

Stolarsky, K.B.

[1977] Power and exponential sums of digital sums related to binomial coefficient parity, *SIAM J. Appl. Math.* **32**, 717–730.

Szegö, G.

[1959] *Orthogonal Polynomials,* revised edition, *Amer. Math. Soc. Coll. Publ.* **23**.

Szekeres, G.

[1953] Some asymptotic formulae in the theory of partitions. II, *Quart. J. Math. Oxford Ser. 2* **4**, 96–111.

[1958] Regular iteration of real and complex functions, *Acta Math.* **100**, 103–258.

[1982] Distribution of labelled trees by diameter, in: *Combinatorial Mathematics X, Proc. 10th Australian Conf. on Combinatorial Mathematics, Lecture Notes in Mathematics,* Vol. 1036 (Springer, Berlin) pp. 392–397.

[1987] Asymptotic distribution of the number and size of parts in unequal partitions, *Bull. Aust. Math. Soc.* **36**, 89–97.

[1990] Asymptotic distribution of partitions by number and size of parts, in: *Number Theory,* Vol. I, eds. K. Györy and G. Halasz, *Colloq. Math. Soc. János Bolyai* **51**, 527–538.

Szpankowski, W.

[1988] The evaluation of an alternative sum with applications to the analysis of some data structures, *Inform. Process. Lett.* **28**, 13–19.

Takacs, L.
[1990] On the number of distinct forests, *SIAM J. Discr. Math.* **3**, 574 581.
[1991] A Bernoulli excursion and its various applications, *Adv. in Appl. Probab.* **23**, 557–585.
Temme, N.M.
[1993] Asymptotic estimates of Stirling numbers, *Studies Appl. Math.* **89**, 233–243.
Titchmarsh, E.C.
[1939] *The Theory of Functions*, 2nd Ed. (Oxford University Press, London).
[1948] *Fourier Integrals*, 2nd Ed. (Oxford University Press, Oxford).
Trjitzinsky, W.J.
[1933a] Analytic theory of linear q-difference equations, *Acta Math.* **61**, 1–38.
[1933b] Analytic theory of linear differential equations, *Acta Math.* **62**, 167–226.
Varadarajan, V.S.
[1991] Meromorphic differential equations, *Expositiones Math.* **9**, 97–188.
Vaughan, R.C.
[1981] *The Hardy–Littlewood Method* (Cambridge University Press, Cambridge).
Vershik, A.M., and C.V. Kerov
[1977] Asymptotics of the Plancherel measure of the symmetric group and a limiting form for Young tableau, *Dokl. Akad. Nauk USSR* **233**, 1024–1027 (in Russian).
Vitter, J., and P. Flajolet
[1990] Analysis of algorithms and data structures, in: *Handbook of Theoretical Computer Science*, Vol. A: *Algorithms and Complexity*, ed. J. Van Leeuwen (North-Holland, Amsterdam) ch. 9, pp. 432–524.
Wasow, W.
[1965] *Asymptotic Expansions for Ordinary Differential Equations* (Wiley, New York).
Wei, W.D., Y.Z. Cai, C.L. Liu and A.M. Odlyzko
[1986] Balloting labelling and personnel assignment, *SIAM J. Algebra Discr. Methods* 7, 150–158.
Whitehead Jr, E.A.
[1979] Four-discordant permutations, *J. Aust. Math. Soc. Ser. A* **28**, 369–377.
Whittaker, E.T., and G.N. Watson
[1927] *A Course of Modern Analysis*, 4th Ed. (Cambridge University Press, Cambridge).
Wilf, H.S.
[1986] The asymptotics of $e^{P(z)}$ and the number of elements of each order in S_n, *Bull. Amer. Math. Soc.* **15**, 228–232.
[1990] *Generatingfunctionology* (Academic Press, New York).
[1993] The asymptotic behavior of the Stirling numbers of the first kind, *J. Combin. Theory A* **64**, 344–349.
Wilf, H.S., and D. Zeilberger
[1990] Rational functions certify combinatorial identities, *J. Amer. Math. Soc.* **3**, 147–158.
[1992] An algorithmic proof theory for hypergeometric (ordinary and "q") multisum/integral identities, *Inventiones Math.* **108**, 575–633.
Wilson, R.
[1953] The coefficient theory of integral functions with dominant exponential parts, *Quart. J. Math. Oxford Ser. 2* **4**, 142–149.
Wimp, J.
[1984] *Computation with Recurrence Relations* (Pitman, Boston).
[1991] Current trends in asymptotics: some problems and some solutions, *J. Comput. Appl. Math.* **35**, 53–79.
Wimp, J., and D. Zeilberger
[1985] Resurrecting the asymptotics of linear recurrences, *J. Math. Anal. Appl.* **111**, 162–176.
Wong, R.
[1989] *Asymptotic Approximations of Integrals* (Academic Press, New York).
Wong, R., and M. Wyman
[1974] The method of Darboux, *J. Approx. Theory* **10**, 159–171.
Wright, E.M.
[1931] Asymptotic partition formulae, I: Plane partitions, *Quart. J. Math. Oxford* **2**, 177–189.

[1932] The coefficients of a certain power series, *J. London Math. Soc.* **7**, 256–262.

[1949] On the coefficients of power series having exponential singularities, *J. London Math. Soc.* **24**, 304–309.

[1958] Partitions of large bipartities, *Amer. J. Math.* **80**, 643–658.

[1959] The asymptotic behavior of the generating functions of partitions of multipartities, *Quart. J. Math. Oxford (2)* **10**, 60–69.

[1961] Partitions into k parts, *Math. Ann.* **142**, 311–316.

[1967] A relationship between two sequences, *Proc. London Math. Soc.* **17**, 296–304, 547–552.

[1970] Asymptotic relations between enumerative functions in graph theory, *Proc. London Math. Soc. (3)* **20**, 558–572.

[1971a] Graphs on unlabelled nodes with a given number of edges, *Acta Math.* **126**, 1–9.

[1971b] The number of strong digraphs, *Bull. London Math. Soc.* **3**, 348–350.

[1971c] Graphs on unlabelled nodes with a large number of edges, *Proc. London Math. Soc. (3)* **28**, 577–594.

Wright, E.M., and B.G. Yates

[1950] The asymptotic expansion of a certain integral, *Quart. J. Math. Oxford* **1**, 41–53.

Wright, R.A., L.B. Richmond, A.M. Odlyzko and B.D. McKay

[1986] Constant time generation of free trees, *SIAM J. Comput.* **15**, 540–548.

Wyman, M.

[1959] The asymptotic behavior of the Laurent coefficients, *Canad. J. Math.* **11**, 534–555.

[1964] The method of Laplace, *Trans. R. Soc. Canada* **2**, 227–256.

Zeilberger, D.

[1979] Solutions of exponential growth to systems of partial differential equations, *J. Differential Equations* **31**, 287–295.

[1980] The algebra of linear partial difference operators and its applications, *SIAM J. Math. Anal.* **11**, 919–932.

[1989] Six etudes in generating functions, *Int. J. Comput. Math.* **29**, 201–215.

[1990] A holonomic approach to special function identities, *J. Comput. Appl. Math.* **32**, 321–368.

CHAPTER 23

Extremal Graph Theory

Béla BOLLOBÁS*

Department of Pure Mathematics, University of Cambridge, 16 Mill Lane, Cambridge, CB2 15B, UK

and

Department of Mathematics, Louisiana State University, Baton Rouge, LA, USA

Contents

* Supported by NSF Grant DMS-8806097.

HANDBOOK OF COMBINATORICS
Edited by R. Graham, M. Grötschel and L. Lovász

Introduction

In extremal graph theory one explores in the relations between various graph invariants like order, size, connectivity, chromatic number, diameter, radius, clique number, minimal and maximal degrees, the circumference, the genus. More generally, one is interested in the values of these invariants ensuring that a graph having a certain property has another given property as well. Let us give two examples. Given a graph F, determine $\mathrm{ex}(n; F)$, the maximal number of edges in a graph of order n that does not contain F, the *forbidden graph*, as a subgraph. Given two properties of graphs, $\mathcal{P}$ and $\mathcal{Q}$, say, a number of graph invariants $f_1, \ldots, f_k$, and a natural number n, determine the set $A(n) = \{(a_1, \ldots, a_k)$: if a graph G of order n with $f_i(G) = a_i$, $i = 1, \ldots, k$, has property $\mathcal{P}$ then it also has property $\mathcal{Q}\}$.

The first of these is the classical extremal problem which, though important, is rather narrow; the second problem, on the other hand, is perhaps too broad a problem to be rightly claimed as a genuine extremal problem, since most problems in graph theory could be formulated in this way. In practice, one stays away from both extremes by considering a problem in graph theory to be an extremal problem if its "natural" formulation asks for some best possible inequalities among various graph invariants. However, in this chapter we shall take a rather narrow view of extremal problems, mostly for lack of space and also because several problems belonging to extremal graph theory are considered in other chapters of this volume, in chapters on Ramsey theory, Hamilton cycles, colouring, connectivity, matching, etc.

In a typical extremal problem, given a property $\mathcal{P}$ and an invariant ϕ for a class $\mathcal{G}$ of graphs, we wish to determine the least value f for which every graph G in $\mathcal{G}$ with $\phi(G) > f$ has property $\mathcal{P}$. The graphs in $\mathcal{G}$ without property $\mathcal{P}$ and satisfying $\phi(G) = f$ are the *extremal graphs* for the problem. More often than not, $\mathcal{G}$ consists of graphs of the same order n, namely $\mathcal{G} = \{G \in \mathcal{H}\colon |G| = n\}$, where $\mathcal{H}$ is a class of graphs, and so f is considered to be a function of n, determined by ϕ and $\mathcal{H}$. This function $f(n)$ is the *extremal function* for the problem.

A short review like this is easily overcrowded with a host of results. In order to avoid this, in section 1 we shall study the classical extremal problem, the problem of forbidden subgraphs, at a leisurely pace, giving some of the simpler proofs. The other sections are considerably shorter and are intended to provide the reader with only glimpses of the topics. Our aim is to give the flavour of the subject rather than overwhelm the reader with results. This review is based mostly on Bollobás (1978a) and an update of that book, Bollobás (1986).

1. Forbidden subgraphs

Let $\mathcal{F} = \{F_1, \ldots, F_k\}$ be a family of graphs of order at most n: the family of *forbidden graphs*. Write $\mathrm{ex}(n; \mathcal{F}) = \mathrm{ex}(n; F_1, \ldots, F_k)$ for the maximal size of a graph of order n containing no forbidden graph F_i, i.e., containing no subgraph isomorphic to a forbidden graph F_i. In this section we shall take $\mathcal{F}$ to be a fixed family, independent of n, and we are mostly interested in the asymptotic value of $\mathrm{ex}(n;\mathcal{F})$ as $n\to\infty$.

1.1. *Turán's theorem and its extensions*

One of the earliest substantial theorems in graph theory is due to Turán (1941) and it concerns the function $\mathrm{ex}(n;K_r)$, where K_r is the complete graph of order r. Turán's theorem was not only the starting point of extremal graph theory but it also signalled the birth of graph theory as an active subject. Although Mantel (1907) proved that $\mathrm{ex}(n;K_3)=\lfloor n^2/4\rfloor$, Turán was the first to study $\mathrm{ex}(n;K_r)$ for all r.

Given $1\leqslant s\leqslant n$, denote by $T_s(n)$ the complete s-partite graph with $\lfloor n/s\rfloor$, $\lfloor (n+1)/s\rfloor,\ldots,\lfloor (n+s-1)/s\rfloor$ vertices in the various classes. Thus $T_s(n)$ is the unique complete s-partite graph of order n whose classes are as equal as possible. Equivalently, it is also the unique s-partite graph of order n whose size is as large as possible. The graph $T_s(n)$ is the *s-partite Turán graph of order n*. Denote the *size*, i.e., the number of edges, of $T_s(n)$ by $t_s(n)$:

$$t_s(n)=\binom{n}{2}-\sum_{i=1}^{s}\binom{n_i}{2}=\sum_{1\leqslant i<j\leqslant s}\left\lfloor\frac{n+i-1}{s}\right\rfloor\left\lfloor\frac{n+j-1}{s}\right\rfloor,$$

where $n_i=\lfloor (n+i-1)/s\rfloor$ is the number of vertices in the ith smallest class. In particular, $t_2(n)=\lfloor n^2/4\rfloor$.

An $(r-1)$-partite graph does not contain a K_r; in particular, $T_{r-1}(n)$ does not contain a K_r. Consequently, $\mathrm{ex}(n;K_r)\geqslant t_{r-1}(n)$. Turán (1941) (see also Turán 1954) proved that, in fact, we have equality, and $T_{r-1}(n)$ is the only extremal graph.

Theorem 1.1.1. *Let* $r\geqslant 2$. *Then* $\mathrm{ex}(n;K_r)=t_{r-1}(n)$ *and* $T_{r-1}(n)$ *is the only extremal graph: it is the only graph of order* n *and size* $t_{r-1}(n)$ *that contains no complete graph of order* r.

Proof. The graph $T_{r-1}(n)$ is a maximal K_r-free graph: it contains no K_r and if we join two vertices belonging to the same class of $T_{r-1}(n)$ then these two vertices, together with $r-2$ vertices, one from each of the other classes, form a K_r. Hence it suffices to prove the second assertion: if G has order n, size $t_{r-1}(n)$, and contains no K_r, then G is (isomorphic to) $T_{r-1}(n)$.

The structure of $T_{r-1}(n)$ is ideal for proving this by induction on n. Indeed, given that we have $t_{r-1}(n)$ edges, the vertices in $T_{r-1}(n)$ have as equal degrees as possible: the minimal degree is $\delta_{r-1}(n)=\lfloor 2t_{r-1}(n)/n\rfloor=n-\lfloor (n+r-2)/(r-1)\rfloor=n-\lceil n/(r-1)\rceil$ and the maximal degree is $\Delta_{r-1}(n)=\lceil 2t_{r-1}(n)/2\rceil=n-\lfloor n/(r-1)\rfloor$. Furthermore, if we delete a vertex x of minimal degree from $T_{r-1}(n)$ then we obtain $T_{r-1}(n-1)$. In particular, $t_{r-1}(n)-\delta_{r-1}(n)=t_{r-1}(n-1)$. Finally, as $\delta_{r-1}(n) = n - 1 - \lfloor (n-1)/(r-1) \rfloor$, the vertex x is joined to all $T_{r-1}(n-1)$ except to the vertices in a smallest class.

Let us see then the proof by induction on n. For $n \leqslant r-1$ there is no prove so let us assume that $n \geqslant r$ and the assertion holds for smaller values of n. Let G be a graph of order n and size $t_{r-1}(n)$ that does not contain a K_r. Let $x \in G$ be a vertex of minimal degree: $d(x) = \delta(G) \leqslant \lfloor 2e(G)/n \rfloor = \lfloor 2t_{r-1}(n)/n \rfloor = \delta_{r-1}(n)$. Set $H = G - x$. Then $e(H) = e(G) - d(x) \geqslant t_{r-1}(n) - \delta_{r-1}(n) = t_{r-1}(n-1)$. Since H contains no K_r, by the induction hypothesis H is $T_{r-1}(n-1)$ and $d(x) = \delta_{r-1}(n)$. The vertex x cannot be joined to $r-1$ vertices in distinct classes of $H = T_{r-1}(n-1)$ because then these r vertices would form a K_r. Consequently $T_{r-1}(n-1)$ has a class, no vertex of which is joined to x. But then this has to be a smallest class and x has to be joined to all the vertices in all the other classes. Therefore G is precisely $T_{r-1}(n)$. □

The proof above is not so much about graphs not containing a complete graph of order r but about the unusual ease with which $T_{r-1}(n)$ can be produced from $T_{r-1}(n-1)$. Let us give a slightly different slant to the proof of the induction step above. Since the degrees of the vertices of $T_{r-1}(n)$ are as equal as possible, given the number of edges, and since $e(G) = t_{r-1}(n)$, there is a vertex x in G with $d(x) \leqslant \delta_{r-1}(n) = \delta(T_{r-1}(n))$. Then, by the induction hypothesis, $H = G - x$ must be $T_{r-1}(n-1)$ and $d(x) = \delta_{r-1}(n)$. If the vertices not joined to x form a (smallest) class of $T_{r-1}(n-1)$ then we are done. Otherwise pick a vertex y in $T_{r-1}(n-1)$ which is not joined to x. Then y has degree $\delta_{r-1}(n)$ in G so, by the induction hypothesis, $G - y$ is also $T_{r-1}(n-1)$. But that is clearly not the case because, for example, $G - y$ contains a K_r.

This version of the proof of the induction step implies the following extension of Theorem 1.1.1.

Theorem 1.1.2. *Let $F_1, \dots, F_k$ be graphs of order at most t, and let s be such that no $T_s(n)$ contains any of the F_i. Suppose $n_0 \geqslant t$ is such that $\mathrm{ex}(n_0; F_1, \dots, F_k) = t_s(n_0)$ and $T_s(n_0)$ is the only extremal graph. Then the same assertion holds for every $n \geqslant n_0$: $\mathrm{ex}(n; F_1, \dots, F_k) = t_s(n)$ and $T_s(n)$ is the only extremal graph.*

If we do not care about the uniqueness of the extremal graph $T_{r-1}(n)$ in Theorem 1.1.1, then all we need for the proof is that every graph of order $n \geqslant r+1$ and size $t_{r-1}(n) + 1$ has minimal degree at most $\delta_{r-1}(n)$. This observation shows that if G is a graph of order n and size $t_{r-1}(n) + 1$ then for every n', $r + 1 \leqslant n' \leqslant n$, the graph G contains a subgraph of order n' and size at least $t_{r-1}(n') + 1$. In particular, as shown by Dirac (1963), every graph of order $n \geqslant r+1$ and size $t_{r-1}(n) + 1$ contains not only a K_r but also a K_{r+1}^-, a complete graph of order $r+1$ from which an edge has been deleted.

This observation can be carried over to greater excess size over $t_s(n)$. A graph G of order $n \geqslant (2q-1)s + 2$ and size $t_s(n) + q$ has minimal degree at most $\delta_s(n)$, so G has a subgraph of order $n - 1$ and size $t_s(n-1) + q$. This implies the following result.

Theorem 1.1.3. *Let $s \geq 2$, $q \geq 1$, $n_0 \geq (2q-1)s+2$ and let $F_1, \ldots, F_k$ be graphs such that* $\mathrm{ex}(n_0; F_1, \ldots, F_k) \leq t_s(n_0) + q$. *Then* $\mathrm{ex}(n; F_1, \ldots, F_k) \leq t_s(n) + q$ *for all* $n \geq n_0$.

Let us return to Turán's Theorem 1.1.1. This result claims that the size of a graph G of order n not containing a K_r is dominated by the size of an $(r-1)$-partite graph H of order n. Erdős (1970) proved that we can guarantee that this domination holds at every vertex: the edges of G can be rearranged and, perhaps, some more edges can be added to the graph in such a way that the resulting graph H is $(r-1)$-partite and every vertex is incident with at least as many edges in H as in G. As so often in mathematics (especially in combinatorics), the achievement is the *discovery* of this beautiful fact: the proof is straightforward.

Theorem 1.1.4. *Let G be a graph not containing a K_r, $r \geq 2$. Then there is an $(r-1)$-partite graph H with vertex set $V(H) = V(G) = V$ such that $d_G(x) \leq d_H(x)$ for every $x \in V$. Furthermore, H can be chosen to satisfy $e(G) < e(H)$, i.e., $d_G(x) < d_H(x)$ for at least one vertex x, unless G is a complete $(r-1)$-partite graph with $r-1$ non-empty classes.*

Proof. We apply induction on r. The assertion is obvious for $r=2$, so we pass to the induction step. Suppose $r>2$ and the assertion holds for smaller values of r. Let $v \in V$ be a vertex of maximal degree in G: $d_G(z) = \Delta(G)$, and let $W = \Gamma(v)$ be the set of neighbours of v. Then $\tilde{G} = G[W]$, the graph induced by W, does not contain a K_{r-1}. Hence, by the induction hypothesis, there is an $(r-2)$-partite graph $\tilde{H}$ with vertex set W such that $d_{\tilde{G}}(w) \leq d_{\tilde{H}}(w)$ for every $w \in W$.

Let us construct an $(r-1)$-partite graph H with vertex set V from $\tilde{H}$ by joining all vertices in $V \backslash W$ to all vertices in W. It is easily seen that $d_G(x) \leq d_H(x)$ for every $x \in V$. Furthermore, it is easily seen that if $\tilde{G}$ is a complete $(r-2)$-partite graph and $d_G(x) = \Delta(G)$ for every $x \in V \backslash W$ then G is a complete $(r-1)$-partite graph. □

Since $T_{r-1}(n)$ is the *unique* $(r-1)$-partite graph of order n and maximal size, Theorem 1.1.4 implies Theorem 1.1.1.

Let us say a few words about a natural extension of the function $\mathrm{ex}(n; \mathcal{F})$. For a graph G and a family $\mathcal{F}$ of graphs, let $\mathrm{ex}(G; \mathcal{F})$ be the maximal number of edges in a subgraph of G that contains no element of $\mathcal{F}$ as a subgraph. Thus, $\mathrm{ex}(n; \mathcal{F}) = \mathrm{ex}(K_n; \mathcal{F})$. It would be unreasonable to expect precise results about the function $\mathrm{ex}(G; \mathcal{F})$ or even $\mathrm{ex}(G; K^r)$ but, somewhat surprisingly, sharp results can be obtained in the case when G is a random graph $G_{n,p}$ (see Bollobás 1985, and chapter 6). Among other results, Babai et al. (1990) proved that, for a fixed value of p, with probability tending to 1, $\mathrm{ex}(G_{n,p}; K^r)$ is the maximal number of edges in an $(r-1)$-partite subgraph of $G_{n,p}$. They also conjectured the following result which was proved, a little later, by Frankl and Pach (1988).

Let us say that a graph has *property* $P(k, l)$ if any k vertices have at most l common neighbours.

Theorem 1.1.5. *Let t, $r \geqslant 2$ be fixed integers, and let $0 < c \leqslant 1 - 1/(r-1)$. Let G be a K_r-free graph with n vertices, having property $P(t, cn)$. Then*

$$e(G) \leqslant c^{1/t}\left(1 - \frac{1}{r-1}\right)^{1-1/t} n^2/2 + o(n^2)\,.$$

As an easy consequence of this result, one finds that $\mathrm{ex}(G_{n,p}; K_r) = p(1 - 1/(r-1))n^2/2 + o(n^2)$ with probability tending to 1.

1.2. The number of complete subgraphs

We know from Turán's theorem that a graph of order greater than $t_{r-1}(n)$ contains at least one K_r, and we know also that it has to contain at least two K_r. Let us go further: given $m > t_{r-1}(n)$, at least how many K_r are in a graph of order n and size m? Even more, if we know that a graph of order n has many K_p subgraphs, what can we say about the minimal number of K_r subgraphs it has to contain?

To formulate this problem precisely, let us introduce some notation. Denote by $k_r(G)$ the number of K_r in a graph G. Thus $k_2(G)$ is just the size of G, the number of edges of G, and Turán's theorem tells us that if G has order n and $k_2(G) > t_{r-1}(n)$ then $k_r(G) \geqslant 1$. For natural numbers $2 \leqslant p < r \leqslant n$ and a real number $x \geqslant 0$ define

$$k_r(k_p^n \geqslant x) = \min\{k_r(G^n) \colon G^n \text{ is a graph of order } n \text{ and } k_p(G^n) \geqslant x\}\,.$$

What can we say about the function $k_r(k_p^n \geqslant x)$? As shown by Bollobás (1976a), this function is also closely connected with the Turán graphs $T_2(n)$, $T_3(n), \ldots$. For simplicity, let us suppress the variable n and put $T_q = T_q(n)$. The graph $G = T_{r-1}$ contains no K_r but it has $k_p(T_{r-1})$ complete graphs of order p, so $k_r(k_p^n \geqslant x) = 0$ for $0 \leqslant x \leqslant k_p(T_{r-1})$.

Let $\psi(x)$ be the *maximal convex function* defined on the interval $k_p(T_{r-1}) \leqslant x \leqslant \binom{n}{p}$ such that

$$\psi(k_p(T_q)) \leqslant k_r(T_q) \tag{1}$$

for $q = r-1, r, \ldots, n$. It is easily seen that, in fact, equality holds in (1) for every q. Also, the Turán graph T_q shows that for $x = k_p(T_q)$ we have

$$k_r(k_p^n \geqslant x) \leqslant \psi(x)\,. \tag{2}$$

It turns out that $\psi(x)$ is actually a lower bound for $k_r(k_p^n \geqslant x)$ for all values of x.

Theorem 1.2.1. *Let $2 \leqslant p < r \leqslant n$. For $k_p(T_{r-1}) \leqslant x \leqslant \binom{n}{p}$ we have*

$$k_r(k_p^n \geqslant x) \geqslant \psi(x)\,.$$

In particular, if a graph of order n has at least as many K_p subgraphs as $T_q(n)$ then it also has at least as many K_r subgraphs as $T_q(n)$. Also, if a graph of order n has more K_p subgraphs than $T_{r-1}(n)$ then it contains a K_r.

The last assertion above was first proved by Erdős (1962) and it was rediscovered by Sauer (1971).

Let us state a weaker but more transparent version of Theorem 1.2.1. The bound on the number of triangles given below was conjectured by Nordhaus and Stewart (1963).

Theorem 1.2.2. (i) *Let $n^2/4 \leq m \leq n^2/3$. Then every graph of order n and size m contains at least $n(4m - n^2)/9$ triangles.*

(ii) *Every graph of order n and size m contains at least $n^{r-2}(2(r-1)m - (r-2)n^2)/r^{r-1}$ copies of K_r.*

The bound above on the minimal number of triangles is fairly good: it is certainly best possible for $n = 3n_0$ and $m = n^2/3 = 3n_0^2$. However, when m is not much greater than $t_2(n) = \lfloor n^2/4 \rfloor$ then the estimate is rather crude. How can we construct a graph of order n and size $m = \lfloor n^2/4 \rfloor + l$ which contains few triangles? For $l < n/2$ we can join a vertex in a larger class of $T_2(n)$ to l vertices of the same class to obtain a graph containing precisely $l\lfloor n/2 \rfloor$ triangles. Erdős (1962) conjectured that we can never do better and proved that this is indeed the case if $l < cn$ for some $c > 0$. This conjecture was proved by Lovász and Simonovits (1976, 1983), who also proved a number of results concerning $k_r(k_2^n \geq x)$, the minimal number of complete r-graphs in a graph of order n, with at least x edges.

Theorem 1.2.3. *For $0 < l < n/2$, a graph with n vertices and $t_2(n) + l$ edges contains at least $l\lfloor n/2 \rfloor$ triangles.*

There are a good many results concerning the covering of graphs by complete subgraphs. The first result in this area was proved by Erdős et al. (1966b); this was sharpened by Bollobás (1976a), Chung (1981) and Győri and Kostochka (1979). The following result was conjectured by Erdős and proved by Pyber (1986).

Theorem 1.2.4. *Let G be a graph with n vertices. Then G and its complement can be covered with at most $\lfloor n^2/4 \rfloor + 2$ complete subgraphs. The graph $T_2(n)$ shows that this bound is best possible.*

A considerable extension of the original theorem of Erdős et al. was conjectured by Winkler, and proved by McGuinness (1994).

Theorem 1.2.5. *If maximal cliques are removed one by one from a graph with n vertices, then the graph will be empty after at most $n^2/4$ steps.*

In fact, Winkler made a stronger conjecture as well, which is still open: if maximal cliques are removed one by one from a graph with n vertices, then the graph will be empty after the sum of the number of vertices in the cliques has reached $n^2/2$.

1.3. Complete bipartite graphs

Let us turn to the analogue of the Turán problem for bipartite graphs. Given natural numbers m, n, s and t, what is the maximal size of an m by n *bipartite graph* not containing a $K(s,t)$, a complete s by t bipartite graph? Denote this maximum by $z(m,n;s,t)$ and put $z(n;t)=z(n,n;t,t)$. Zarankiewicz (1951) asked this question for $s=t=3$ and $m=n=4$, 5, 6 and the general problem has also become known as the *problem of Zarankiewicz*. The similarity with Turán's problem is, unfortunately, only superficial: for the general function $z(m,n;s,t)$ there is no beautiful extremal graph and we are far from being able to determine even the order of $z(n;t)$ for a fixed (but large) value of t.

It is worth reformulating the Zarankiewicz problem in terms of 0–1 matrices. At most how many 1s can a 0–1 matrix of m rows and n columns contain if it has no s by t submatrix all whose entries are 1s?

The following rather trivial lemma is just about the most one can say about the general function $z(m,n;s,t)$. As, trivially, $z(m,n;1,t)=m(t-1)$ for $1\leq t\leq n$, it is sufficient to consider the case $2\leq s\leq m$, $2\leq t\leq n$.

Lemma 1.3.1. *Let m, n, s, t, r and k be integers, $2\leq s\leq m$, $2\leq t\leq n$, $0\leq r\leq m$, and let G be an m by n bipartite graph of size $z=my=km+r$ without a $K(s,t)$. Then*

$$m\binom{y}{t}\leq(m-r)\binom{k}{t}+r\binom{k+1}{t}\leq(s-1)\binom{n}{t}. \tag{1}$$

Proof. Let (V_1,V_2) be the bipartition of G and let $V_1=\{x_1,\ldots,x_m\}$, $d(x_i)=d_i$. Let us call a set $\{xy_1,xy_2,\ldots,xy_t\}$ of t edges of G incident with the same vertex x a *claw*; furthermore, x is the *centre* of the claw and the t-set $\{y_1,\ldots,y_t\}$ is the *base*.

The graph G has $\sum_{i=1}^{m}\binom{d_i}{t}$ claws since there are $\binom{d_i}{t}$ claws with centre x_i. On the other hand, each t-subset of V_2 is the base of at most $s-1$ claws since G contains no $K(s,t)$. Therefore G has at most $(s-1)\binom{n}{t}$ claws and so

$$\sum_{i=1}^{m}\binom{d_i}{t}\leq(s-1)\binom{n}{t}. \tag{2}$$

Since $\sum_{i=1}^{m}d_i=z=km+r$, $0\leq r<m$, and $\binom{u}{t}$ is a convex function of u for $u\geq t$, inequality (2) implies (1). □

Theorem 1.3.2. *Let m, n, s, t be natural numbers, $2\leq s\leq m$, $2\leq t\leq n$. Then*

$$z(m,n;s,t)<(s-1)^{1/t}(n-t+1)m^{1-1/t}+(t-1)m\,.$$

Proof. Let G be an extremal graph for $z(m, n; s, t)$. Set $y = z(m, n; s, t)/m$. Then, since $y < n$, by Lemma 1 we have

$$m(y - (t-1))^t < (s-1)(n-(t-1))^t . \quad \square$$

For a fixed value of $t \geq 2$, Theorem 1.3.2 implies that

$$z(n; t) \leq (t-1)^{1/t} n^{2-1/t} + O(n) \tag{3}$$

and it is conjectured that (3) is essentially best possible. To be precise, it is conjectured that

$$\lim_{n\to\infty} z(n; t)/n^{2-1/t} = c_t > 0 \tag{4}$$

for every $t \geq 2$. So far, the only value of t for which (4) is known to hold is $t = 2$. In fact, Kővári et al. (1954) and Reiman (1958) determined $z(n; 2)$ for infinitely many values of n, but there is no $t \geq 3$ for which $z(n; t)$ is known for infinitely many values of n.

Theorem 1.3.3. (i) $z(n; 2) \leq (n/2)\{1 + \sqrt{4n-3}\}$ *for all* $n \geq 2$.

(ii) *Let q be a prime power and let $n = q^2 + q + 1$. Then*

$$z(n; 2) = \frac{n}{2}\{1 + \sqrt{4n-3}\} = (q-1)(q^2+q+1) .$$

(iii) $\lim_{n\to\infty} z(n; 2)/n^{3/2} = 1$.

Proof. (i) Let G be an extremal graph for $z(n; 2)$ and let the notation be as in the proof of Lemma 1.3.1. By inequality (2),

$$\binom{n}{2} \geq \sum_{i=1}^{n} \binom{d_i}{2}$$

$$n^2 - n \geq \sum_{i=1}^{n} d_i^2 - \sum_{i=1}^{n} d_i \geq \Big(\sum_{i=1}^{n} d_i\Big)^2/n - \sum_{i=1}^{n} d_i = z^2/n - z .$$

This implies the required inequality.

(ii) From the proof of part (i) we see that equality holds in (i) if and only if (1) every vertex in G has the same degree d, (2) for every two vertices in V_2 there is precisely one vertex in V_1 joined to both, and (3) for every two vertices in V_2 there is precisely one vertex in V_1 joined to both. This means that the graph G can be considered as a finite projective plane: V_1 is the set of points, V_2 is the set of lines and $x \in V_1$ is joined to $y \in V_2$ iff the point x is incident with the line y. Now if q is a prime power then there is a projective plane of order q, that is with $n = q^2 + q + 1$ points and lines.

(iii) Since for every sufficiently large natural number n, there is a prime between $n - n^{2/3}/10$ and n, the assertion follows from (i) and (ii). $\square$

A somewhat weaker form of conjecture (4) is that $\varliminf_{n\to\infty} z(n;t)/n^{2-1/t} > 0$. In addition to $t=2$, this is known for $t=3$. Brown (1966) proved that $\varliminf_{n\to\infty} z(n;3)/n^{2-1/3} \geq 1$ by making use of the 3-dimensional affine space AG(3, p) over the finite field of order p. However, for a general $t \geq 4$ all we know is that

$$\varliminf_{n\to\infty} z(n;t)/n^{2-2/(t+1)} \geq 1-(t!)^{-2}\,. \tag{5}$$

This is proved by making use of random graphs (see Bollobás 1979, p. 127). The gap between the upper bound, $n^{2-1/t}$, and the lower bound, $n^{2-2/(t+1)}$, is alarmingly large; as stated above, it is very likely that the upper bound gives the correct value.

The functions $\mathrm{ex}(n;K(s,t))$ and $z(n,n;s,t)$ are intimately connected; in particular, for fixed values of s and t they have the same order. It is easily seen that

$$2\,\mathrm{ex}(n;K(s,t)) \leq z(n,n;s,t) \leq \mathrm{ex}(2n;K(s,t))\,. \tag{6}$$

Indeed, given a graph G of order n and size $m=\mathrm{ex}(n;K(s,t))$, construct an n by n bipartite graph H as follows. Take two disjoint copies of $V(G)$, say V_1 and V_2, and join $x' \in V_1$ to $y'' \in V_2$ iff $xy \in E(G)$, where x and y are the vertices in $V(G)$ corresponding to x' and y''. Then H has $2m$ edges and contains no $K(s,t)$ (and no $K(t,s)$, for that matter) so the first inequality in (6) holds. The second inequality is trivial.

Combining inequality (6) with Theorem 1.3.2, and noting the analogue of (5), we have the following assertion.

Theorem 1.3.4. *If* $2 \leq s < n$ *then*

$$\begin{aligned}\tfrac{1}{2}(1-(s!)^{-2})n^{2-2/(s+1)} &\leq \mathrm{ex}(n;K(s,s))\\ &\leq \tfrac{1}{2}(s-1)^{1/s}(n-s+1)n^{1-1/s}+\tfrac{1}{2}(s-1)n\\ &< n^{2-1/s}+\frac{s-1}{2}n\,.\end{aligned}$$

As (6) holds and we do not know the order of $z(n,n;t,t)$ for $t \geq 4$, neither do we know the order of $\mathrm{ex}(n;K(s,s))$ for $s \geq 4$. However, we do know that $\mathrm{ex}(n;K(2,2))$ has order $n^{3/2}$ and $\mathrm{ex}(n;K(3,3))$ has order $n^{5/3}$. In the case of $K(2,2)$ we can do considerably better. As in the problem of determining $\mathrm{ex}(n;K(2,2))$ we do not care where the classes of $K(2,2)$ are, it is more natural to write C_4 instead of $K(2,2)$, indicating that $K(2,2)$ is just a *4-cycle* or *quadrilateral*.

Inequality (5) and Theorem 1.3.3 (ii) imply that

$$\mathrm{ex}(n;C_4) \leq \frac{n}{4}\{1+\sqrt{4n-3}\}\,. \tag{7}$$

Erdős et al. (1966a) noticed that certain graphs constructed by Erdős and Rényi

(1962) show that (6) is asymptotically best possible. The same assertion was proved independently by Brown (1966).

Theorem 1.3.5. *Let q be a prime power. Then*

$$\tfrac{1}{2}q(q+1)^2 \le \mathrm{ex}(q^2+q+1; C_4) \le \tfrac{1}{2}q(q+1)^2 + \frac{q+1}{2}. \tag{8}$$

Furthermore,

$$\lim_{n\to\infty} \mathrm{ex}(n; C_4)/n^{3/2} = \tfrac{1}{2}. \tag{9}$$

Proof. The second inequality is precisely inequality (6) for $n = q^2 + q + 1$. Let us prove the first inequality by describing the graph G_q constructed by Erdős and Rényi (1962).

The vertex set $V(G_0)$ is the set of $q^2 + q + 1$ points of the finite projective plane PG$(2, q)$ over the finite field of order q. A point is joined to all the points on its *polar* with respect to the conic $x^2 + y^2 + z^2 = 0$. Thus two points (a, b, c) and (α, β, γ) are joined iff $a\alpha + b\beta + c\gamma = 0$. Then a point not on the conic is joined to $q + 1$ points, i.e., to all the lines on its polar, while each of the $q + 1$ points on the conic is joined to q points, namely to the points on its polar except itself. Hence G_q has $\frac{1}{2}\{q^2(q+1) + (q+1)q\} = \frac{1}{2}q(q+1)^2$ edges.

The graph G_q does not contain a quadrilateral since any two lines meet in exactly one point so every vertex is determined by any two of its neighbours.

Relation (9) follows as Theorem 1.3.3 (iii). □

The bounds in (7) are tantalizingly close. The only reason why the graph G_q is not ideal for the problem is that it has *absolute points*, i.e., points lying on their polars. These $q + 1$ points are joined to only q points, instead of $q + 1$, as all the others. If we could avoid these absolute points by choosing a more suitable polarity then we would achieve the upper bound in (7). However, this is not to be: Baer (1946) proved that every polarity of a finite projective plane of order q has at least $q + 1$ absolute points. Thus the Erdős–Rényi graph G_q cannot be made to have more edges by choosing a different polarity.

In view of this fact it is not too surprising that the way to improve (8) is to reduce the upper bound. This was achieved by Füredi (1983) (see also the remarks at the end of that paper) who thereby determined ex$(n; C_4)$ for infinitely many values of n.

Theorem 1.3.6. *For every natural number q we have*

$$\mathrm{ex}(q^2+q+1; C_4) \le \tfrac{1}{2}q(q+1)^2.$$

In particular, if q is a prime power then

$$\mathrm{ex}(q^2+q+1; C_4) = \tfrac{1}{2}q(q+1)^2.$$

What happens if we forbid not only C_4 but C_5 as well? The projective graph in Theorem 1.3.3 (ii) contains no C_4, and as it is bipartite, it contains no C_5 either. Hence if $n = 2(q^2 + q + 1)$ for some prime power q then $\mathrm{ex}(n; C_4, C_5) \geqslant (q-1)(q^2 + q + 1)$, so $\mathrm{ex}(n; C_4, C_5) \geqslant (n/2)^{3/2} + \mathrm{o}(n^{3/2})$ for all n. Erdős and Simonovits (1982) proved that this inequality is, in fact, an equality.

Theorem 1.3.7. $\mathrm{ex}(n; C_4, C_5) = (n/2)^{3/2} + \mathrm{o}(n^{3/2})$.

It would be of interest to decide whether $\mathrm{ex}(n; C_4, C_5) = (q-1)(q^2 + q + 1)$ if q is a prime power and $n = 2(q^2 + q + 1)$.

1.4. *The fundamental theorem of extremal graph theory*

For $r \geqslant 3$, the Turán graph $T_{r-1}(n)$ has $t_{r-1}(n) = (r - 2/2(r-1))n^2 + \mathrm{O}(n)$ edges and contains no K_r. On the other hand, every graph of order n and size $t_{r-1}(n) + 1$ has a K_r, in fact, several K_r. Furthermore, Theorem 1.2.2 implies that if $0 < \varepsilon < 1/2r(r-1)$ then every graph of order n and size $((r-2)/2(r-1) + \varepsilon)n^2$ contains at least $(2(r-1)\varepsilon/r^{r-1})n^r$ copies of K_r. Thus there is a sudden jump when the size reaches $t_{r-1}(n)$.

Although this sudden jump is quite startling, Erdős and Stone (1946) proved that a considerably more important change takes place when the size becomes significantly greater than $t_{r-1}(n)$. This result, which deserves to be called the *fundamental theorem of extremal graph theory*, states that for every $r \geqslant 3$ and $\varepsilon > 0$, there is a function $s = s(n)$ such that $s(n) \to \infty$ as $n \to \infty$, and every graph of order n and size $(((r-2)/2(r-1)) + \varepsilon)n^2$ contains a $K_r(s) = K(s, s, \ldots, s) = T_r(rs)$, a complete r-partite graph with s vertices in each of the classes. Thus we not only get a complete r-partite graph with one vertex in each class, as claimed by Turán's theorem, but we can guarantee even a complete r-partite graph with $s(n)$ vertices in each class, where $s(n) \to \infty$ as $n \to \infty$.

The assertion above does make sense for $r = 2$ as well although in that case Turán's theorem is completely trivial: every graph of order n and size at least εn^2, $0 < \varepsilon < \frac{1}{2}$, contains a complete bipartite graph with at least $s(n)$ vertices in each class, where $s(n) \to \infty$ as $n \to \infty$. This assertion is immediate from Theorem 1.3.4: if $0 < \varepsilon < \frac{1}{2}$ and $0 < c < \log 1/2\varepsilon$ are fixed then the assertion is true with $s(n) = \lceil c \log n \rceil$, provided n is sufficiently large.

Let us state then the fundamental theorem of extremal graph theory, proved by Erdős and Stone (1946).

Theorem 1.4.1. *Let $r \geqslant 2$ and $\varepsilon > 0$ be fixed. Then there is a function $s = s(n)$, with $\lim_{n \to \infty} s(n) = \infty$, such that every graph of order n and size at least $((r-2)/2(r-1) + \varepsilon)n^2$ contains a $K_r(s)$.*

As we are interested in the growth of $s(n)$, let us introduce the following notation. For $r \geq 2$ and $0 < \varepsilon < 1/2(r-1)$ define

$$s_{r,\varepsilon}(n) = \min\left\{t\colon \text{every graph of order } n \text{ and size at least } \left(\frac{r-2}{2(r-1)} + \varepsilon\right)n^2 \text{ contains a } K_r(t)\right\}.$$

Erdős and Stone (1946) proved that $s_{r,\varepsilon}(n) \geq (l_{r-1}(n))^{1/2}$ if n is sufficiently large, where $l_{r-1}(n)$ is the $r-1$ times iterated logarithm of n. Furthermore, Erdős and Stone conjectured that the order of $s_{r,\varepsilon}(n)$ is $l_{r-1}(n)$. Later Erdős (1967) announced that $s_{r,\varepsilon}(n) > c(\log n)^{1/(r-1)}$ for some constant $c > 0$ and sufficiently large n.

Rather unexpectedly, $s_{r,\varepsilon}(n)$ turns out to be much larger than these lower bounds. The true order of $s_{r,\varepsilon}(n)$ was determined by Bollobás and Erdős (1973).

Theorem 1.4.2. *Let $r \geq 2$ and $0 < \varepsilon < 1/2(r-1)$. Then there are positive constants $c_1 = c_1(r, \varepsilon)$ and $c_2 = c_2(r, \varepsilon)$ such that*

$$c_1 \log n < s_{r,\varepsilon}(n) < c_2 \log n\,. \tag{1}$$

In particular, every graph of order n and size at least $((r-2)/2(r-1) + \varepsilon)n^2$ contains a complete r-patite graph with at least $c_1 \log n$ vertices in each class.

How do c_1 and c_2 depend on r and ε? As pointed out by Bollobás and Erdős (1973), the constant c_2 can be chosen to be $5/\log(1/\varepsilon)$, provided n is sufficiently large. This can be seen by a simple application of random graphs. What about c_1? Improving inequality (1), Bollobás et al. (1976) proved that one can take $c_1 = c/r \log(1/\varepsilon)$ for some absolute constant $c > 0$, provided n is sufficiently large. Finally, Chvátal and Szemerédi (1981) showed that this is true without the factor r.

Theorem 1.4.3. *There is an absolute constant $c > 0$ such that*

$$\frac{c}{\log(1/\varepsilon)} \log n < s_{r,\varepsilon}(n) < \frac{5}{\log(1/\varepsilon)} \log n$$

if $r \geq 2$, $0 < \varepsilon < 1/2(r-1)$ and n is sufficiently large.

First we shall sketch a proof of Theorem 1.4.2 and then we shall return to Theorem 1.4.3. As we remarked above, the upper bound in (1) is very easy: it follows from a straightforward application of random graphs. To prove the lower bound, we shall need the following lemma.

Lemma 1.4.4. *Let G be a graph of order n that contains no $K_{r+1}(s)$ but contains a $K_r(q)$, say $\tilde{K}$. Then G has at most*

$$((r-1)q + s)n + 2qn^{1-1/s}$$

edges joining $\tilde{K}$ to $G - \tilde{K}$.

Proof. As in the proof of Lemma 1.3.1, we define a *claw with centre* $x \in G - \tilde{K}$ as the set of r edges incident with x such that precisely s of these edges join x to each of the r classes of $\tilde{K}$. It is easily checked that if $x \in G - \tilde{K}$ is joined to $(r-1)q + d$ vertices in $\tilde{K}$ then there are at least $\binom{q}{s}^{r-1}\binom{d}{s}$ claws with centre x. Hence if there are $(r-1)qn + D > (r-1)qn + sn$ edges joining $G - \tilde{K}$ to $\tilde{K}$ then there are at least $n\binom{q}{s}^{r-1}\binom{D/n}{s}$ claws in G.

Since G contains no $K_{r+1}(s)$, there are at most $s-1$ claws with the same *base*, the same set of vertices joined to the centre. As there are $\binom{q}{s}^r$ possible bases, G contains at most $(s-1)\binom{q}{s}^r$ claws. Consequently,

$$n\binom{D/n}{s} \leqslant (s-q)\binom{q}{s}.$$

Hence

$$D \leqslant n^{1-1/s}(s-1)^{1/s}q \leqslant 2n^{1-1/s}q\,,$$

proving the lemma. □

Armed with this lemma, we shall prove the main part of Theorem 1.4.2, the lower bound on $s_{r,\varepsilon}(n)$. To be precise, we shall prove the following assertion.

Theorem 1.4.2′. *Let $r \geqslant 2$, $0 < \varepsilon < 1/2(r-1)$ and $0 < \gamma_r < (r-1)!\varepsilon^{r-1}/\log(8/\varepsilon)$. Then if n is sufficiently large, every graph of order n and size at least*

$$\left(\frac{r-2}{2(r-1)} + \varepsilon\right)n^2$$

contains a $K_r(s)$ where $s = \lfloor \gamma_r \log n \rfloor$.

Proof. Let us add to Theorem 1.4.2 a trivial assertion concerning the case $r = 1$: for $\varepsilon > 0$, every graph of sufficiently large order contains a $K_1(s)$ for $s = \lfloor \gamma_1 \log n \rfloor$ where $\gamma_1 = 2/\varepsilon$.

Suppose then that the result is true for $r \geqslant 1$ but fails for $r+1$: there is a constant γ'_{r+1}, $0 < \gamma'_{r+1} < r!\varepsilon^r/\log(8/\varepsilon)$, such that for every n_0 there is a graph G_1 of order $n_1 \geqslant n_0$ and size at least $((r-1)/2r + \varepsilon)n_1^2$ without a $K_{r+1}(s_1)$, where $s_1 = \lfloor \gamma'_{r+1} \log n_1 \rfloor$. Such a graph G_1 has average degree $((r-1)/r + 2\varepsilon)n_1$ so it contains a subgraph G with $n \geqslant \frac{1}{2}\varepsilon n_1$ vertices and minimal degree at least $((r-1)/2r + \frac{3}{2}\varepsilon)n$. Let $\gamma'_{r+1} < \gamma_{r+1} < \varepsilon r \gamma_r < r!\varepsilon^r/\log(8/\varepsilon)$. Then, if n is sufficiently large (and that is the case if n_0 is sufficiently large), the graph G contains no $K_{r+1}(s)$, where $s = \lfloor \gamma_{r+1} \log n \rfloor$. However, it does contain a $K_r(q)$, say $\tilde{K}$, where $q = \lfloor \gamma_r \log n \rfloor$. By Lemma 1.4.4 there are at most $((r-1)q + s)n + 2qn^{1-1/s}$ edges joining $\tilde{K}$ to $G - \tilde{K}$, so some vertex of $\tilde{K}$ has degree at most

$$rq + \{((r-1)q + s)n + 2qn^{1-1/s}\}/rq\,.$$

Hence

$$\left(\frac{r-1}{r}+\frac{3}{2}\varepsilon\right)n \leq \delta(G) \leq \frac{r-1}{r}n + rq + \frac{sn}{rq} + \frac{2}{r}ns^{1-1/s}.$$

This inequality cannot hold if n is large enough since then $rq < \frac{1}{4}\varepsilon n$, $s/rq < \varepsilon$ and $(2/r)n^{-1/s} < \frac{1}{4}\varepsilon$. This contradiction completes the proof. □

The proof Theorem 1.4.3, given by Chvátal and Szemerédi (1981), is based on a deep and important lemma due to Szemerédi (1978). This result, to be stated below as Theorem 1.4.5 and usually called the *uniform density lemma* or *regularity lemma*, was one of the main tools in the proof of Szemerédi's (1975) theorem, one of the most difficult results in combinatorics, stating that every sequence of integers with positive upper density contains arbitrarily long arithmetic progressions.

For a graph G, and disjoint sets $U, W \subset V(G)$, denote by $e(U, W)$ the number of $U - W$ edges. The *density* of the edges between U and W is

$$d(U, W) = \frac{e(U, W)}{|U||W|}.$$

The pair (U, W) is *ε-uniform* or *ε-regular* if

$$|d(U', W') - d(U, W)| < \varepsilon$$

whenever $U' \subset U$, $W' \subset W$, $|U'| > \varepsilon|U|$ and $|W'| > \varepsilon|W|$.

Theorem 1.4.5. *Given $\varepsilon > 0$ and an integer m, there is an $M = M(\varepsilon, m)$ such that the vertices of every graph of order at least m can be partitioned into classes V_0, $V_1, \ldots, V_k$, where $m \leq k \leq M$, such that $|V_0| \leq |V_1| = |V_2| = \cdots = |V_k|$ and all but at most εk^2 of the pairs (V_i, V_j), $1 \leq i < j \leq k$, are ε-uniform.*

The following two immediate consequences of Theorem 1.4.1 show why the result is called the fundamental theorem of extremal graph theory. In the spirit of the notation used above, for a graph G and a set $U \subset V(G)$ define the *density* $d(U)$ of the subgraph $G[U]$ spanned by U as

$$d(U) = e(G[U])\Big/\binom{u}{2},$$

where $u = |U|$. Thus if U spans a complete graph then $d(U) = 1$, if U consists of independent vertices then $d(U) = 0$.

Let G be an infinite graph. Define the *upper density of* G to be

$$\bar{d}(G) = \sup\{\alpha\colon \text{for every } m > 0 \text{ there is a finite set } U \text{ satisfying } |U| > m \text{ and } d(U) > \alpha\}.$$

Putting it another way, if $\beta > \bar{d}(G)$ then there is an $m > 0$ such that whenever U

has at least m vertices then $d(U) < \beta$, and $\bar{d}(G)$ is the smallest such number. Clearly, if G is the empty graph then $\bar{d}(G) = 0$, also, if G contains arbitrarily large complete graphs then $\bar{d}(G) = 1$. What are the possible values of the upper densities? It is rather natural to expect the closed interval to be the set of possible upper densities. Surprisingly, this is not the case.

Theorem 1.4.6. *The set of upper densities of infinite graphs is* $\{1, 0, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \ldots\}$.

Proof. Suppose $\bar{d}(G) > 1 - 1/r + \varepsilon$ for some $r \in \mathbb{N}$ and $\varepsilon > 0$. Then G contains a sequence of subgraphs, say $G_1, G_2, \ldots$ such that G_i has order n_i and size at least $((r-1)/r + \frac{3}{4}\varepsilon)\binom{n_i}{2} > ((r-1)/2r + \frac{1}{3}\varepsilon)n_i^2$, and $n_i \to \infty$. By Theorem 1.4.1, each G_i contains a $K_{r+1}(s_i)$, where $s_i \to \infty$. Now $d(K_{r+1}(s_i)) > r/(r+1)$ and the order of $K_{r+1}(s_i)$ tends to ∞, so $\bar{d}(G) \geq r/(r+1)$. □

The other immediate consequence of Theorem 1.4.1 concerns the approximate value of $\mathrm{ex}(n; F_1, F_2, \ldots, F_k)$. As observed by Erdős and Simonovits (1966), Theorem 1.4.1 implies that $\lim_{n\to\infty} \mathrm{ex}(n; F_1, \ldots, F_k)/\binom{n}{2}$ is a very simple function of the family $\{F_1, \ldots, F_k\}$.

Theorem 1.4.7. *Let* $F_1, \ldots, F_k$ *be fixed non-empty graphs. Set* $r = \min_i \chi(F_i) - 1$, *i.e., let* $r + 1$ *be the smallest chromatic number of an* F_i. *Then*

$$\lim_{n\to\infty} \mathrm{ex}(n; F_1, \ldots, F_k) \Big/ \binom{n}{2} = 1 - \frac{1}{r}\,.$$

Proof. We may assume that $\chi(F_1) = r + 1$. The graph $T_r(n)$ contains no F_i so $\mathrm{ex}(n; F_1, \ldots, F_k) \geq t_r(n) = (1 - 1/r)\binom{n}{2} + (n)$. Hence $\varliminf_{n\to\infty} \mathrm{ex}(n; F_1, \ldots, F_k)/\binom{n}{2} \geq 1 - 1/r$.

On the other hand, if $\varepsilon > 0$ and n is sufficiently large, then by Theorem 1.4.1 every graph G of order n and size at least $(1 - 1/r + \varepsilon)\binom{n}{2}$ contains a $K_{r+1}(s)$, where $s > |F_1|$. But then $K_{r+1}(s)$ contains F_1 and, therefore, so does G. As this holds for every $\varepsilon > 0$, we have

$$\varlimsup_{n\to\infty} \mathrm{ex}(n; F_1, \ldots, F_k) \Big/ \binom{n}{2} \leq 1 - \frac{1}{r}\,. \qquad \square$$

Although Theorem 1.4.7 is just an immediate corollary of Theorem 1.4.1, at the first sight it is, nevertheless, very surprising: the crude order of $\mathrm{ex}(n; F_1, \ldots, F_k)$ depends only on the minimal chromatic number of the F_i. In particular, the asymptotic value of $\mathrm{ex}(n; F_1, \ldots, F_k)$ is easily determined if no F_i is bipartite. Of course, this leaves several questions unanswered. What is the error term $\phi(n)$ in $\mathrm{ex}(n; F_1, \ldots, F_k) = ((r-1)/2r)n^2 + \phi(n)$? What is the asymptotic value of $\mathrm{ex}(n; F_1, \ldots, F_k)$ when some F_i is bipartite? We know from section 1.3 that we are far from being able to answer the third question for an arbitrary family, since we do not even know the asymptotic value of $\mathrm{ex}(n; K_{4,4})$, say, but we shall discuss the first two questions in section 1.5.

Let us note an easy application of Theorem 1.4.7, giving the rough solution of a seemingly intractable problem.

Theorem 1.4.8. *Let $\mathcal{F}$ be the family of graphs of order p and size q. Let* $r = \min\{s$: $t_{s+1}(p) \geqslant q\}$. *Then*

$$\lim_{n\to\infty} \mathrm{ex}(n; \mathcal{F})/n^2 = \frac{r-1}{2r}\,.$$

Proof. Note that $\min\{\chi(F)\colon\ F \in \mathcal{F}\} = r + 1$. □

To conclude this section, we state a weak form of Theorem 1.4.6, as it leads to some deep questions concerning r-graphs, i.e., r-uniform hypergraphs. Given $r \geqslant 2$ and $0 \leqslant \alpha < 1$, we say that α is a *jump value for r-graphs* if there is a $\beta > \alpha$ such that if $\varepsilon > 0$, $m \geqslant r$ and $n \geqslant n(\alpha, \varepsilon, m)$ then every r-graph with $n \geqslant n(\alpha, \varepsilon, m)$ vertices and at least $\alpha\binom{n}{r}$ hyperedges contains a subgraph with m vertices and at least $\beta\binom{m}{r}$ hyperedges. Note that α is a jump value for graphs if for some $\delta > 0$ the interval $(\alpha, \alpha + \delta)$ contains no upper density of an infinite graph. Hence the following result is immediate from either Theorem 1.4.1 or Theorem 1.4.6.

Theorem 1.4.9. *Every $0 \leqslant \alpha < 1$ is a jump value for graphs.*

Erdős posed the problem of deciding whether the same is true for r-graphs. The problem was open for several years and was eventually solved by Frankl and Rődl (1984).

Theorem 1.4.10. *Let $r \geqslant 3$ and $s > 2r$ be natural numbers. Then $1 - s^{1-r}$ is not a jump value for r-graphs.*

This beautiful and difficult problem leaves open a number of important questions. In particular, it would be interesting to determine the set of jump values for r-graphs and the set of upper densities for r-graphs.

1.5. The structure of extremal graphs

For a family $\mathcal{F} = \{F_1, \ldots, F_k\}$ of forbidden graphs, denote by $\mathrm{EX}(n; \mathcal{F}) = \mathrm{EX}(n; F_1, \ldots, F_k)$ the *set of extremal graphs of order n*. Thus a graph G belongs to $\mathrm{EX}(n; \mathcal{F})$ iff G has order n, size $\mathrm{ex}(n; \mathcal{F})$ and contains no forbidden graph, i.e., no member of $\mathcal{F}$. Turán's theorem, Theorem 1.1.1, tells us that $\mathrm{EX}(n; K_r) = \{T_{r-1}(n)\}$ for all r and n, $2 \leqslant r \leqslant n$. For a general family $\mathcal{F}$, Theorem 1.4.6, an immediate consequence of the Erdős–Stone theorem, the fundamental theorem of extremal graph theory, gives us the rough order of $\mathrm{ex}(n; \mathcal{F})$. But what is the more precise order of $\mathrm{ex}(n; \mathcal{F})$ for a general family $\mathcal{F}$ and what do extremal graphs look like?

These questions were answered, surprisingly precisely, by Erdős and

Simonovits (1966) and by Simonovits (1968); simpler proofs of the results can be found in Bollobás (1978a, pp. 339–345). Here we shall state only of the results.

Theorem 1.5.1. *Let F be a graph with $\chi(F) = r + 1 \geq 3$, and for $n = 1, 2, \ldots$ let G^n be a graph of order n and size $(1 - 1/r + \mathrm{o}(1))\binom{n}{2}$ not containing F. Then the following assertions hold.*

(i) *There is a $K(p_1, p_2, \ldots, p_r)$, $\sum_{i=1}^{r} p_i = n$, $p_i = (1 + \mathrm{o}(1))n/r$, that can be obtained from G^n by adding and subtracting $\mathrm{o}(n^2)$ edges.*

(ii) *G^n contains an r-partite graph of size $(1 - 1/r + \mathrm{o}(1))\binom{n}{2}$.*

(iii) *G^n contains an r-partite graph of minimal degree $(1 - 1/r + \mathrm{o}(1))n$.*

The result above claims that if a graph G^n not containing F has *about* as many edges as the Turán graph $T_r(n)$, which *trivially* fails to contain F, then G^n is *very close* to the graph $T_r(n)$. For an extremal graph, considerably more is true.

Theorem 1.5.2. *Let $\mathcal{F} = \{F_1, \ldots, F_k\}$ be a fixed family of graphs, let $r + 1 = \min_i \chi(F_i) \geq 2$ and suppose that F_1 has an $(r + 1)$-colouring in which one of the colour classes contains t vertices. Let $G^n \in \mathrm{EX}(n; \mathcal{F})$. Then, as $n \to \infty$,*

$$e(G^n) = \mathrm{ex}(n; \mathcal{F}) = \left(1 - \frac{1}{r}\right)\binom{n}{2} + O(n^{2-1/t}),$$

$$\delta(G^n) = \left(1 - \frac{1}{r} + \mathrm{o}(1)\right)n,$$

the vertices of G can be partitioned into r classes such that each vertex is joined to at most as many vertices in its own class as in any other class, and for every $\varepsilon > 0$, there are at most $c(\varepsilon, \mathcal{F})$ vertices joined to at least εn vertices of the same class. Furthermore, there are $\mathrm{O}(n^{2-1/t})$ edges joining vertices belonging to the same class, and each class has $n/r + \mathrm{O}(n^{2-1/t})$ vertices.

This result gives us a very good hold on extremal graphs. In fact, the function $\mathrm{O}(n^{2-1/t})$ can be replaced by $\mathrm{O}(\mathrm{ex}(n; K(s, t))$ where s and t are fixed. In particular, we have the following better bound on $\mathrm{ex}(n; F)$ in terms of $\mathrm{ex}(m; F_0)$ for some bipartite graph F_0.

Theorem 1.5.3. *Let $F = F_0 + K_{r-1}(u)$ where F_0 is a bipartite graph. Then*

$$\mathrm{ex}(n; F) \leq \left(1 - \frac{1}{r}\right)\binom{n}{2} + (r + \mathrm{o}(1))\mathrm{ex}\left(\left\lfloor \frac{n}{r} \right\rfloor; F_0\right).$$

As an illustration of the power of Theorem 1.5.2, let us present a beautiful theorem of Simonovits (1968) giving a complete solution to the forbidden subgraph problem for sK_{r+1}, i.e., for s disjoint copies of K_{r+1}, provided n is sufficiently large.

What is a likely candidate for an extremal graph for sK_{r+1}? If we add $t - 1$ vertices to the Turán graph $T = T_r(n - t + 1)$ and join these vertices to each other and to the vertices of T then the obtained graph, $K_{s-1} + T_r(n-t+1)$, has quite a few more (about $(t-1)n/r$ more) edges than $T_r(n)$, the extremal graph for one copy of K_{r+1}, and still fails to contain s disjoint copies of K_{r+1}. Indeed, every K_{r+1} in $K_{t-1} + T_r(n-t+1)$ must contain at least one of the $t-1$ vertices of K_{t-1}. The following theorem of Simonovits (1968) shows that our hunch is essentially correct.

Theorem 1.5.4. *Let r and s be fixed natural numbers, $r \geq 2$. If n is sufficiently large then $K_{t-1} + T_r(n-t+1)$ is the unique extremal graph for tK_{r+1}.*

Proof. Let us apply induction on t. The case $t = 1$ is precisely Turán's theorem, so let us pass to the induction step.

Let $G = G^n$ be an extremal graph of order n for tK_{r+1}, and consider the partition $V = V_1 \cup V_2 \cup \cdots \cup V_r$ guaranteed by Theorem 1.5.2. Set $\varepsilon = 1/4tr$. Let us distinguish two cases.

Case (i) *Some vertex x is joined to at least εn vertices in its own class.* Let W_i be a set of $m = \lceil \varepsilon n \rceil$ neighbours of x in V_i. By Theorem 1.5.2 the r-partite subgraph of G spanned by $W_1 \cup W_2 \cup \cdots \cup W_r$ has $(1 - 1/r + o(1))r^2m^2/2$ edges so, rather trivially (or by Theorem 1.4.1, if we wish to conclude it instantly), it contains a $K_r(s)$ for $s = t(r+1)$, provided n is sufficiently large. But then $G - x$ cannot contain a $(t-1)K_{r+1}$ since any $(t-1)K_{r+1}$ could be extended to a tK_{r+1} so we are done by the induction hypothesis.

Case (ii) *Every vertex is joined to at most εn vertices in its own class.* In this case, our aim is to arrive at a contradiction. As $\delta(G) \geq (1 - 1/r + o(1))n$, we may assume that every vertex is joined to all but at most $2\varepsilon n$ vertices in the other classes. As in case (i), this implies that for every pair $\{x, y\}$ of vertices in the same class, in particular, for every edge xy joining vertices in the same class, the graph G contains a $K_{r-1}(s)$ for $s = t(r+1)$ such that both x and y are joined to all vertices of this $K_{r-1}(s)$. But then this implies that the graph H obtained from G by deleting all edges joining different classes contains at most $t-1$ independent edges.

Recall that the maximal degree of H is at most εn. Let $\{x_1y_1, \ldots, x_ky_k\}$, $k \leq t-1$, be a maximal set of independent edges in H. Since every edge of H meets the set $\{x_1, y_1, x_2, y_2, \ldots, x_k, y_k\}$, we have $e(H) \leq 2k\varepsilon n < 2t\varepsilon n$. But then

$$t_r(n) + \frac{n}{r} - 1 < \mathrm{ex}(n; tK_{r+1}) = e(G) < t_r(n) + 2t\varepsilon n ,$$

contradicting the choice of ε, provided n is sufficiently large. □

A good many substantial general results concerning the structure of graphs in $\mathrm{EX}(n; F)$ were proved by Simonovits (1968, 1974).

Another result based on Theorem 1.5.2, a theorem of Bollobás et al. (1978), shows the surprisingly great difference one edge can make.

Let q be a prime power and let $n = q^2 + q + 1$. Let G be the graph obtained from $K(n, n)$ by placing an Erdős–Rényi graph G_q, described in the proof of

Theorem 1.3.5, in each of the classes. Thus

$$e(G) = n^2 + q(q+1)^2 = q^4 + 3q^3 + 5q^2 + 3q + 1 .$$

As G_q has maximal degree $q+1$ and contains no $C_4 = K(2,2)$, the maximal t for which G contains a $K(2,2,t)$ is precisely $q + 1 \sim \sqrt{n}$. *One more* edge guarantees the existence of a $K(2,2,\lfloor \gamma n \rfloor)$ where $\gamma > 0$ is an absolute constant.

Theorem 1.5.5. *There is a constant q_0 such that if $q \geqslant q_0$ is a prime power and $n = q^2 + q + 1$ then*

$$\mathrm{ex}(2n; K(2,2,q+2)) = n^2 + q(q+1)^2 .$$

Furthermore, every graph of order $2n$ and size $n^2 + q(q+1)^2 + 1$ contains a $K(2,2,t)$ with $t \geqslant 10^{-3}n$.

To conclude this section, we present a theorem of Erdős and Simonovits (1983). This result is related to Theorem 1.2.3: it concerns the number of $\mathcal{F}$-subgraphs of a graph with n vertices and *substantially* more than $\mathrm{ex}(n; \mathcal{F})$ edges. Similarly to the notation $k_r(G)$ used earlier, given a family $\mathcal{F}$ of graphs, denote by $k_{\mathcal{F}}(G)$ the number of subgraphs of a graph G isomorphic to elements of $\mathcal{F}$. Thus $\mathrm{ex}(n; \mathcal{F}) = \max\{e(G)$: G has n vertices and $k_{\mathcal{F}}(G) = 0\}$. The following result is a special case of a theorem of Erdős and Simonovits (1983), proved for hypergraphs.

Theorem 1.5.6. *Let $\mathcal{F}$ be a finite family of graphs, with each $F \in \mathcal{F}$ having at least t vertices. Then for every constant $c > 0$ there is a constant $c' > 0$ such that if G is a graph with n vertices and at least* $\mathrm{ex}(n; \mathcal{F}) + cn^2$ edges then $k_{\mathcal{F}}(G) \geqslant c'n^t$.

1.6. The asymptotic number of graphs without forbidden subgraphs

Given a forbidden graph F, denote by $f(n; F)$ the number of graphs on $[n] = \{1, 2, \ldots, n\}$ not containing F. What can we say about $f(n; F)$ as $n \to \infty$? As always, we are particularly interested in the case $F = K_r$. Extending earlier results of Erdős et al. (1976), Kolaitis et al. (1987) proved the following beautiful and sharp theorem.

Theorem 1.6.1. *For $r \geqslant 3$, $f(n; K_r)$ is asymptotic to the number of $(r-1)$-partite graphs on $[n]$. In particular,*

$$\begin{aligned} f(n; K_r) &= 2^{\{(r-1)/2(r-1) + o(1)\}n^2} \\ &= 2^{(1+o(1))\mathrm{ex}(n;\, K_r)} . \end{aligned}$$

As we shall see, Theorem 1.6.1 and a simple application of Szemerédi's uniformity lemma (Theorem 1.4.5) enable one to determine the asymptotic value of $\log f(n; F)$ for every F of chromatic number at least 3.

Let us start with a trivial lower bound for $f(n; F)$. *If a graph* G on $[n]$ does not contain our forbidden graph F, then no subgraph of G contains F and so

$$f(n; F) \geq 2^{e(G)} .$$

Since G can be chosen to have $\mathrm{ex}(n; F)$ edges, we find that

$$f(n; F) \geq 2^{\mathrm{ex}(n; F)} .$$

Erdős et al. (1986) showed that this trivial bound is not far from being best possible. The key to this result is the following property of ε-uniform and fairly dense pairs (see Theorem 1.4.5 and the paragraph preceding it).

Lemma 1.6.2. *Let* $f \geq 1$, $r \geq 2$ *and* $0 < \varepsilon < (r-1)^{-1/2}$, *and let* $V_1, \ldots, V_r$ *be disjoint subsets of the vertex set* $V(G)$ *of a graph* G *with* $(1-\varepsilon)\varepsilon^{f-1}|V_i| \geq 1$, $i = 1, \ldots, r$. *Suppose that each pair* (V_i, V_j) *is* ε^f*-uniform with density at least* $\varepsilon + \varepsilon^2$. *Then* G *contains every r-partite graph on f vertices.*

Proof. We apply induction on f. As for $f = 1$ there is nothing to prove we turn to the induction step: we assume that $f \geq 2$ and that the lemma holds for smaller values of f.

For every i, $2 \leq i \leq r$, the set V_1 has at most $\varepsilon^f|V_1|$ vertices joined to fewer than $(d(V_1, V_i) - \varepsilon^f)|V_i| \geq \varepsilon|V_i|$ vertices of V_i. Hence there are at least $(1-(r-1)\varepsilon^f)|V_1| > 0$ vertices in V_1, each of which is joined to at least $\varepsilon|V_i|$ vertices in V_i, $i = 2, \ldots, r$. Let x_1 be such a vertex and set $W_1 = V_1 \backslash \{x_1\}$ and $W_i = \Gamma(x_1) \cap V_i$, $i = 2, \ldots, r$. Then $|W_1| = |V_1| - 1 \geq \varepsilon|V_1|$ and $|W_i| \geq \varepsilon|V_i|$ for $i = 2, \ldots, r$. Hence, the sets $W_1, \ldots, W_r$ satisfy the conditions of the lemma with f replaced by $f-1$. As x_1 is joined to all vertices in $\bigcup_{i=2}^r W_i$, we are done by the induction hypothesis. □

We know from section 1.5 that the structure of an extremal graph for F is rather close to the structure of an extremal graph for K_r, where $r = \chi(F)$. The following theorem of Erdős et al. (1986) claims that *any* graph not containing F can be turned into a graph not containing K_r by the deletion of a few edges.

Theorem 1.6.3. *For every* $\varepsilon > 0$ *and graph* F *there is a constant* $n_0 = n_0(\varepsilon, F)$ *with the following property: Let* G *be a graph of order* $n \geq n_0$ *not containing* F *as a subgraph. Then* G *contains a set* E' *of at most* εn^2 *edges such that* $G \backslash E'$ *contains no* K_r, *where* $r = \chi(F)$.

Proof. We may assume that $r \geq 3$ and $\varepsilon < 2/(r-1)$. Set $f = |F|$, $m = \lceil 3/\varepsilon \rceil$ and $\varepsilon_0 = \varepsilon/4$. Let $M = M(\varepsilon_0^f, m)$ be the constant guaranteed by Szemerédi's uniformity lemma (Theorem 1.4.5).

We claim that $n_0 = n_0(\varepsilon, F) = \lceil (M+1)/\varepsilon_0^f \rceil$ will do. Indeed, let G be a graph of order $n \geq n_0$ not containing F. By Theorem 1.4.5 there is a partition $\bigcup_{i=0}^k V_i$ of $V(G)$ into disjoint sets such that $m \leq k \leq M$, $|V_0| \leq |V_1| = |V_2| = \cdots = |V_k|$, and all but at most $\varepsilon_0^f k^2$ of the pairs (V_i, V_j), $1 \leq i < j \leq k$, are ε_0^f-uniform. Let E' be the union of the following sets of edges:

(1) the edges meeting V_0,

(2) the edges joining two vertices of V_i, $i = 1, \ldots, r$,

(3) the edges joining V_i to V_j for every pair (V_i, V_j) which is not ε_0^f-uniform,

(4) the edges joining V_i to V_j for every pair (V_i, V_j) of density less than $\varepsilon_0 + \varepsilon_0^2$.

By Lemma 1.6.2, the graph $G \backslash E'$ contains no K_r since otherwise it would contain F as well. Hence all we have to check is that E' is small enough. This is indeed the case:

$$\begin{aligned}
|E'| &\leq \frac{n^2}{k+1} + k\binom{n/k}{2} + \varepsilon_0^f k^2 (n/k)^2 + (\varepsilon_0 + \varepsilon_0^2)\binom{n}{2} \\
&< n^2\left\{\frac{1}{k} + \frac{1}{2k} + \varepsilon_0^f + \varepsilon_0\right\} \\
&\leq n^2\left\{\frac{3}{2m} + \frac{\varepsilon}{2}\right\} \leq \varepsilon n^2 . \qquad \square
\end{aligned}$$

From here it is a short step to the theorem of Erdős et al. (1976) concerning $f(n; F)$.

Theorem 1.6.4. *Let F be a graph with $r = \chi(F) \geq 3$. Then*

$$f(n; F) = 2^{(1+o(1))\mathrm{ex}(n;F)} = 2^{((r-2)/2(r-1)+o(1))n^2} .$$

Proof. Theorem 1.6.3 implies that if $\varepsilon > 0$ and n is sufficiently large then

$$f(n; F) \leq f(n; K_r)\binom{\binom{n}{2}}{\varepsilon n^2} \leq f(n; K_r)(2/\varepsilon)^{\varepsilon n^2} .$$

Hence, by Theorem 1.6.1,

$$f(n; F) \leq 2^{(1+o(1))\mathrm{ex}(N;K_r)+o(n^2)} = 2^{(1+o(1))\mathrm{ex}(n;F)} . \qquad \square$$

It is easily seen that Theorems 1.6.3 and 1.6.4 hold for *families* of forbidden graphs. Thus if $\mathscr{F} = \{F_1, \ldots, F_k\}$, with

$$\min_{1 \leq i \leq k} \chi(F_i) \geq 3 ,$$

then, with the obvious definition,

$$f(n; \mathscr{F}) = f(n; F_1, \ldots, F_k) = 2^{(1+o(1))\mathrm{ex}(n;\mathscr{F})} .$$

It is interesting to formulate the last assertion in terms of monotone properties. A *property* $\mathscr{P}$ of graphs is an infinite class of (finite) graphs which is closed under isomorphism. A property $\mathscr{P}$ is said to be *monotone* if every subgraph of every member of $\mathscr{P}$ is also in $\mathscr{P}$, and it is *hereditary* if every *induced* subgraph of every member of $\mathscr{P}$ is also in $\mathscr{P}$. Thus every monotone property is also hereditary;

furthermore, the intersection of a family of monotone hereditary properties is monotone, and the intersection of hereditary properties is hereditary.

Monotone properties are characterized by forbidden subgraphs. Indeed, given a family $\mathcal{F}$ of finite graphs, let $\mathcal{P}_{\mathcal{F}}$ be the class of graphs having no subgraph isomorphic to a member of $\mathcal{F}$. If $\mathcal{P}_{\mathcal{F}}$ is infinite then it is a monotone property; conversely, every monotone property is obtained in this way.

A monotone property is *principal* if it is obtained by forbidding a simple graph. Clearly, every property is the intersection of a (possibly infinite) family of principal properties.

Let us write $\mathcal{P}^n$ for the set of graphs in $\mathcal{P}$ with vertex set $[n]$. Thus $f(n; \mathcal{F}) = |\mathcal{P}^n_{\mathcal{F}}|$. The remarks above concerning $f(n; \mathcal{F})$ have the following reformulation.

Theorem 1.6.5. *Let $\mathcal{P}_1, \mathcal{P}_2, \ldots$ be monotone properties and set $\mathcal{P} = \bigcap \mathcal{P}_k$. Then*

$$|\mathcal{P}^n| = 2^{o(n^2)}|\mathcal{P}^n_k|$$

for some k. In particular,

$$|\mathcal{P}^n| = 2^{o(n^2)}|\mathcal{Q}^n|$$

for some principal monotone property $\mathcal{Q}$ containing $\mathcal{P}$.

Returning to $f(n; F)$, let us note that it is not known whether Theorem 1.6.4 holds for every bipartite F as well. In fact, it is not even known whether Theorem 1.6.4 holds for a 4-cycle C_4. Since, by Theorem 1.3.5, $\mathrm{ex}(n; C_4) \sim \frac{1}{2}n^{3/2}$, one would like to show that

$$f(n; C_4) = 2^{(1/2+o(1))n^{3/2}} .$$

While the right-hand side is a (trivial) lower bound for $f(n; C_4)$, the best upper bound, due to Kleitman and Winston (1980), is only $2^{cn^{3/2}}$, with c about 1.08.

1.7. The asymptotic number of graphs without forbidden induced subgraphs

Recently Prömel and Steger studied the structure and number of graphs without *induced* forbidden subgraphs. Given a graph F, let $f^*(n; F)$ be the number of graphs on $[n]$ containing no induced subgraph isomorphic to F (briefly, containing no induced F).

At least how large is $f^*(n; F)$? Suppose that there are integers k and l such that no k-partite graph, in which l of the classes have been replaced by complete graphs, contains an induced F. Then, clearly,

$$f^*(n; F) \geq 2^{((k-1)/2k+o(1))n^2} ,$$

since the classes can be chosen to be almost equal and the edges between the classes can be freely chosen.

Prömel and Steger (1992, 1993a,b) proved that this simple lower bound is essentially best possible. Let $\tau(F)$ be the maximal integer r such that for $k = r - 1$ there is an l as above. This somewhat convoluted definition is explained by the fact that $\tau(F)$ is something like the chromatic number $\chi(F)$, which is the maximal integer r such that for $k = r - 1$ no k-partite graph contains F. So the following result, whose proof is based on a generalization of Szemerédi's uniformity lemma to hypergraphs, is the exact analogue of Theorem 1.6.4.

Theorem 1.7.1. *Let F be a graph with $r = \tau(F) \geq 3$. Then*

$$f^*(n; F) = 2^{((r-2)/2(r-1)+o(1))n^2}\,.$$

For the case $F = C_4$, Prömel and Steger (1991) proved much more precise results. It is easily seen that $\tau(C_4) = 3$. Indeed, if $V(G)$ is the disjoint union of the sets V_1 and V_2, with $G[V_1]$ complete and V_2 an independent set (such graphs are known as *split graphs*), then G does not contain an induced C_4. Hence, by Theorem 1.7.1, we have $f^*(n; C_4) = 2^{(1/4+o(1))n^2}$. In fact, considerably more is true.

Theorem 1.7.2. (i) *Almost every graph containing no C_4 is a split graph: $f^*(n; C_4)$ is asymptotic to the number of split graphs on $[n]$.*

(ii) *There are positive constants c_1 and c_2 such that*

$$f^*(n; C_4) \sim c_j(2^{n^2/4+n})/n^{1/2}\,,$$

where $j \equiv n \pmod 2$.

What happens if we forbid a family $\mathcal{F} = \{F_1, F_2, \ldots\}$ of finite graphs as induced subgraphs? Rather surprisingly, unlike the case of forbidden subgraphs, forbidding a *family* $\mathcal{F}$ induced subgraphs is very different from forbidding just one of them. Let $\mathcal{P} = \mathcal{P}_{\mathcal{F}^*}$ be the class of graphs containing an element of $\mathcal{F}$ as an *induced* subgraph. If $\mathcal{P}_{\mathcal{F}^*}$ is infinite then it is a hereditary property; conversely, every hereditary property is obtained in this way.

The growth of $|\mathcal{P}^n|$ for a hereditary property $\mathcal{P}$ depends on the colouring number $r(\mathcal{P})$ of $\mathcal{P}$, defined somewhat similarly to $\tau(F)$. An *(r,s)-colouring* of a graph H is a map $\psi: V(H) \to [r]$ such that $H[\psi^{-1}(i)]$ is complete for $1 \leq i \leq s$ and is empty for $s + 1 \leq i \leq r$. Thus s of the colour classes induce complete graphs and $r - s$ of them induce empty graphs. The *colouring number* $r(\mathcal{P})$ of a property $\mathcal{P}$ of graphs is the maximal r for which there is an s, $0 \leq s \leq r$ such that every (r,s)-colourable graph has property $\mathcal{P}$. Equivalently, $r(\mathcal{P}_{\mathcal{F}^*}) = \max\{r$: for some s, $0 \leq s \leq r$, no $F \in \mathcal{F}$ is (r,s)-colourable$\}$. Note that

$$r(\mathcal{P}_{\mathcal{F}^*}) \geq \inf_{F \in \mathcal{F}} \{\tau(\mathcal{F}) - 1\}\,,$$

with equality if $\mathcal{F} = \{F\}$ but, in general, the inequality may be strict.

Alekseev (1993) and Bollobás and Thomason (1994b) determined the asymptotic size of $\mathcal{P}^n$ for a hereditary property, thereby extending Theorems 1.6.4 and 1.7.1, concerning principal properties.

Theorem 1.7.3. *Let $\mathcal{P}$ be a hereditary property of graphs and let $\mathcal{P}^n$ be the set of graphs in $\mathcal{P}$ with vertex set $[n]$. Then*

$$|\mathcal{P}^n| = 2^{(1-1/r+o(1))n^2/2},$$

where $r = r(\mathcal{P})$ is the colouring number of $\mathcal{P}$.

This result implies that the analogue of Theorem 1.6.5 does not hold for hereditary properties: the intersection of two hereditary properties may be substantially smaller than either of the properties. For example, if $\mathcal{F}_1 = \{K_4\}$, $\mathcal{F}_2 = \{C_7\}$, $\mathcal{P}_i = \mathcal{P}_{\mathcal{F}_i}$, $i = 1, 2$, and $\mathcal{P} = \mathcal{P}_1 \cap \mathcal{P}_2$ then

$$|\mathcal{P}_i^n| = 2^{(1+o(1))n^2/3}$$

for $i = 1, 2$, but

$$|\mathcal{P}^n| = 2^{(1+o(1))n^2/4}.$$

In conclusion, let us note that the analogous problem for uniform hypergraphs is unsolved. If $\mathcal{P}$ is a property of k-graphs (k-uniform hypergraphs) then, as implied by some results of Alekseev (1982) and Bollobás and Thomason (1994a),

$$|\mathcal{P}^n| = 2^{(c+o(1))\binom{n}{k}}$$

for some constant c. However, for $r \geqslant 3$ the possible values for c are not known.

2. Cycles

In section 1 we discussed the forbidden subgraph problem for a *fixed family* of forbidden graphs $\mathcal{F}$ and found this problem to be fairly well understood, provided $\mathcal{F}$ contains no bipartite graph. What can we say about graphs of order n not containing any member of a family $\mathcal{F}_n$ of forbidden graphs, where $\mathcal{F}_n$ *depends on* n? The most frequently studied and best understood case of this problem is when $\mathcal{F}_n$ consists of cycles. In this section we shall discuss some of the results concerning this problem.

2.1. Hamilton cycles

What values of various graph parameters ensure that a graph has a Hamilton cycle? Let us start with the number of edges ensuring a Hamilton cycle: what is $\mathrm{ex}(n; C_n)$? Since a Hamiltonian graph has minimal degree at least 2, every graph of order n and size $\mathrm{ex}(n; C_n) + 1$ must have minimal degree at least 2. It is immediate that the minimal number of edges ensuring that a graph of order n has minimal degree at least 2 is $\binom{n-1}{2}+2$: adding a vertex x to K_{n-1} and joining x to one vertex of K_{n-1} we obtain the unique graph of order n, size $\binom{n-1}{2}+1$, and minimal degree at most 1 (and so precisely 1). A moment's thought shows that the Hamilton cycle problem has the same solution: $\text{ex}(n; C_n)=\binom{n-1}{2}+2$, with the same extremal graph.

Although this seems somewhat disappointing, all it shows that the size in itself is not very effective in forcing a Hamilton cycle. The minimal degree is considerably better. (Contrast this with the remarks following Theorem 1.1.1 in the previous section.) Dirac (1952) proved that a graph of order n and minimal degree at least $n/2$ is Hamiltonian; the graph $K(\lfloor (n-1)/2\rfloor, \lfloor (n+1)/2\rfloor)$ shows that the result is best possible. This theorem of Dirac started the search for various degree conditions that, coupled with some other conditions, like a bound on the connectedness, imply that the graph is Hamiltonian.

As shown by Ore (1960), Dirac's theorem is implied by the following simple lemma, essentially due to Dirac.

Lemma 2.1.1. *Let x_1 and x_n be non-adjacent vertices in a graph G of order n such that $d(x_1)+d(x_n)\geqslant n$. Then G is Hamiltonian iff $G+x_1x_n$ is Hamiltonian.*

Proof. Suppose there is a Hamilton cycle in $G+x_1x_n$. If this cycle does not contain x_1x_n then G is Hamiltonian so we are done. Otherwise G contains a Hamilton path $x_1x_2\cdots x_n$. Since $d(x_1)+d(x_n)\geqslant n$, there is an index i, $2<i<n$, such that x_1 is joined to x_i and x_n is joined to x_{i-1}. But then $x_ix_1x_2\cdots x_{i-1}x_nx_{n-1}\cdots x_i$ is a Hamilton cycle. □

Thus if a graph G is not Hamiltonian and x, y are non-adjacent vertices such that $d(x)+d(y)\geqslant n$ then $G'=G+xy$ is not Hamiltonian either. Of course, if in G' we can find non-adjacent vertices x', y' such that $d'(x')+d'(y')\geqslant n$, where d' denotes the degree in G', then $G''=G'+x'y'$ is not Hamiltonian either, and so on. This led Bondy and Chvátal (1976) to introduce the k-closure of a graph. The *k-closure* $C_k(G)$ of a graph G is the minimal graph H containing G such that for any two non-adjacent vertices x, y of H we have $d_H(x)+d_H(y)\leqslant k-1$. In other words, $C_k(G)$ is the unique graph obtained from G by successively joining all vertices the sum of whose degrees is at least k. Call a property P of graphs *k-stable* if whenever x, y are non-adjacent vertices of G such that $d(x)+d(y)\geqslant k$, and $G+xy$ has property P then so does G. By definition, if P is k-stable and $C_k(G)$ has P then G has P.

Lemma 2.1.1 states precisely that the property of being Hamiltonian (for graphs of order n) is n-stable. (In fact, the proof of Lemma 2.1.1 shows that the property of containing a cycle of length at least k is also n-stable; and it is easily seen that the property of containing a path of length at least l is $(n-1)$-stable.) Thus if $C_n(G)$ is Hamiltonian so is G. In particular, Lemma 1.1.1 implies Dirac's theorem, from whose proof the lemma was distilled.

Theorem 2.1.2. *Let G be a graph of order $n \geqslant 3$ and minimal degree at least $n/2$. Then G is Hamiltonian.*

Proof. Note that $C_n(G)$ is the complete graph K_n. Since K_n is Hamiltonian, so is G. □

The closure operation enables one to prove the theorem of Las Vergnas (1971) for the existence of a Hamilton cycle.

Theorem 2.1.3. *Let G be a graph with vertex set $\{x_1, x_2, \ldots, x_n\}$. Suppose there are no indices i and j such that $x_i x_j$ is not an edge, $d(x_i) + d(x_j) \leqslant n - 1$, $d(x_i) \leqslant i$, $d(x_j) \leqslant j - 1$ and $j \geqslant \max\{i + 1, n - i\}$. Then G is Hamiltonian.*

As an immediate consequence of this result, one obtains Chvátal's (1972) theorem answering a very natural extremal question concerning Hamilton cycles: what sequences d_1, $d_2, \ldots, d_n$ guarantee that if the ith vertex of a graph G of order n has degree at least d_i then G is Hamiltonian? By Dirac's theorem, $\lceil n/2 \rceil$, $\lceil n/2 \rceil, \ldots, \lceil n/2 \rceil$ is such a sequence.

Theorem 2.1.4. (i) *Let $d_1 \leqslant d_2 \leqslant \cdots \leqslant d_n$ be the degree sequence of a graph of order $n \geqslant 3$. Suppose*

$$d_k \leqslant k < \frac{n}{2} \text{ implies } d_{n-k} \geqslant n - k . \tag{1}$$

Then if G has vertex set $\{x_1, x_2, \ldots, x_n\}$ and $d(x_i) \geqslant d_i$ for every i, then G is Hamiltonian.

(ii) *If $(d_k)_1^n$ is the degree sequence of a graph and* (1) *fails then there is a non-Hamiltonian graph with vertex set $\{x_1, x_2, \ldots, x_n\}$ such that $d(x_i) \geqslant d_i$ for every i.*

Analogous results hold for Hamilton paths: if $C_{n-1}(G)$ has a Hamilton path then so does G, and condition (1) gets replaced by the condition that $d_k \leqslant k - 1 < \frac{1}{2}(n - 1)$ implies that $d_{n+1-k} \geqslant n - k$.

There are numerous other sufficient conditions for a graph to be Hamiltonian that do not demand that the vertices have very large degrees. The first notable result of this kind was proved by Nash-Williams (1971). Let us write $\alpha(G)$ for the *independence* (or *stability*) number of a graph G, i.e., for the maximal cardinality of an independent set of vertices.

Theorem 2.1.5. *Let G be a 2-connected graph of order n and minimal degree $\delta(G) \geqslant (n + 2)/3$. If $\delta(G) \geqslant \alpha(G)$ then G is Hamiltonian.*

In proving Theorem 2.1.5, Nash-Williams made use of the following important lemma.

Lemma 2.1.6. *Let C be a longest cycle in a non-Hamiltonian graph G with n vertices. If $G - C$ has a component with at least 2 vertices then $\delta(G) \leqslant (n+1)/3$.*

This lemma has several extensions, including those by Jackson (1980) and Jung (1984).

Häggkvist (1980, 1989) proved the following deep and useful characterization of Hamiltonian graphs of fairly large minimal degree.

Theorem 2.1.7. *Every 2-connected non-Hamiltonian graph with n vertices and minimal degree $\delta \geqslant \frac{8}{17}(n-1)$ contains a set S of $m \geqslant 3\delta - n + 2 > \frac{7}{17}n$ vertices such that in the graph $G - S$ the vertex set cannot be covered by m paths.*

Note that in Theorem 2.1.4 one allows $d(x_i)$ to be strictly greater than d_i. As the following beautiful theorem of Jackson (1980) shows, if we demand that the graph is 2-connected and every vertex has degree *precisely* d, then a rather small value of d guarantees that the graph is Hamiltonian. The proof of this theorem is based on Jackson's extension of Lemma 2.1.6.

Theorem 2.1.8. *Let G be a 2-connected d-regular graph of order n. If $d \geqslant \frac{1}{3}n$ then G is Hamiltonian.*

The Petersen graph shows that, as stated, Theorem 2.1.8 is best possible, at least for $d = 3$. It is easily seen that it is close to being best possible for every $d \geqslant 3$.

What happens if our graph is not only 2-connected but also k-connected for some $k \geqslant 3$? At first sight it seems likely that a considerably smaller degree of regularity will suffice to imply that the graph is Hamiltonian. In particular, as conjectured by Bollobás (1978a, p. 167, Conjecture 36), it seems likely that if G is a d-regular k-connected graph with n vertices and $d \geqslant n/(k+1)$ then G is Hamiltonian. Jackson and Jung showed that this is false for $k \geqslant 4$.

The examples indicate that for a fixed value of k, k-connectedness is hardly any more use in finding Hamilton cycles in regular graphs than 3-connectedness. However, the conjecture may well be true for $k = 3$: if G is a 3-connected d-regular graph with n vertices and $d \geqslant n/4$ then G is Hamiltonian. This was conjectured by Häggkvist as well.

Recently Li Hao (1989a) took the first step towards proving this conjecture by showing that if we demand 3-connectedness then the degree of regularity can be allowed to drop substantially below the $n/3$ bound in Theorem 2.1.8.

Theorem 2.1.9. *Let G be a 3-connected d-regular graph of order n. If $d \geqslant \frac{7}{22}n$ then G is Hamiltonian.*

Note that Theorem 2.1.5 is another extension of Theorem 2.1.1. The following rather simple result in the vein of Theorem 2.1.5 is due to Chvátal and Erdős (1972).

Theorem 2.1.10. *Suppose G has at least three vertices and it is $\alpha(G)$-connected. Then it is Hamiltonian.*

Proof. Let $k = \alpha(G)$. Then $k \geq 2$ so G has a longest cycle C. Then $|C| \geq \delta(G) + 1 \geq k + 1$. Assume that C is not a Hamilton cycle, i.e., there is a vertex $x \in G - C$. Since G is k-connected, there are k independent paths from x to C, i.e., there are $x - x_i$ paths ($i = 1, \ldots, k$) such that any two of them have only the vertex x in common, and any one of them has only the vertex x_i on C.

Giving C some orientation, let x_i^+ be the successor of x_i on C for $i = 1, \ldots, k$. Then, since C is a longest cycle, the set $S = \{x, x_1, x_2, \ldots, x_k\}$ is an independent set, contradicting our assumption that $\alpha(G) \leq k$. Hence C is a Hamilton cycle. □

Given a set S of vertices of a graph G, denote by $N(S)$ the set of neighbours of S: $N(S) = \{x \in G: xy \in E(G)$ for some $y \in S\}$. Fraisse (1986) proved the following essentially best possible condition for a k-connected graph to be Hamiltonian.

Theorem 2.1.11. *Let G be a k-connected graph of order n. Suppose that $|N(S)| > k(n-1)/(k+1)$ whenever S is an independent set of k vertices. Then G is Hamiltonian.*

The following graph constructed by Skupien (1979) shows that Theorem 2.1.11 is close to being best possible: let $n = (k+1)q + k$ and let G be obtained from the vertex-disjoint union of K_k and $k+1$ copies of K_q, by joining each vertex of K_k to every other vertex. Then G is a k-connected non-Hamiltonian graph of order n, in which any k independent vertices have $n - k - q = kq = k(n-k)/(k+1)$ neighbours.

Recently Häggkvist (1989) proved the following substantial extension of Theorem 2.1.5.

Theorem 2.1.12. *Let G be a non-Hamiltonian 2-connected graph of order n, independence number $\alpha \leq (n+1)/2$ and minimal degree $\delta \geq (n+2)/3$. Then, for every k, $1 \leq k \leq \delta + 1$, there exists an independent set S of k vertices such that*

$$|N(S)| \leq \max\{\alpha - 1, n - 2\delta + k - 2\}\ .$$

A consequence of Theorem 2.1.12 is that if G is a 2-connected non-Hamiltonian graph of order n with minimal degree $\delta \geq (n+2)/3$ then it contains an independent set of at least $(n+14)/6$ vertices with at most $(n-1)/2$ neighbours in total.

2.2. Edge-disjoint Hamilton cycles

Suppose the conditions on some set of graph parameters imply that our graph must contain a Hamilton cycle. Does our graph have to have many Hamilton cycles? Does it have to have many edge-disjoint Hamilton cycles? The following striking theorem of Nash-Williams (1971), whose proof is based on Theorem 2.1.6, shows that this is the case if the parameter is the minimal degree. To be precise, Nash-Williams proved the following substantial extension of Dirac's theorem, Theorem 2.1.2.

Theorem 2.2.1. *Let G be a graph of order n and minimal degree at least $n/2$. Then G contains a set of $\lfloor 5(n+10)/224 \rfloor$ edge-disjoint Hamilton cycles.*

Once again, if we demand that our graph be regular then we can guarantee considerably more edge-disjoint Hamilton cycles. Jackson (1979) made use of his Theorem 2.1.8 to deduce the following result.

Theorem 2.2.2. *Let G be a d-regular graph of order $n \geqslant 14$. If $d \geqslant (n-1)/2$ then G contains a set of $\lfloor (n-1)/2 \rfloor$ edge-disjoint Hamilton cycles.*

Theorem 2.2.1 is rather far from being best possible. In the case when the minimal degree is a little larger than $n/2$, Häggkvist (1990) proved the following deep results that are essentially best possible.

Theorem 2.2.3. *Let $\lambda > \frac{1}{2}$. If n is sufficiently large and G is a graph of order n and minimal degree at least λn, then G has a set of $\lfloor n/8 \rfloor$ edge-disjoint Hamilton cycles.*

Theorem 2.2.4. *Let $\lambda > \frac{1}{2}$. If n is sufficiently large and G is a d-regular graph of order n, where d is an even integer not less than λn, then G has a Hamilton decomposition, i.e., the edge set of G can be partitioned into $d/2$ Hamilton cycles.*

To see that, in some sense, Häggkvist's theorem 2.2.3 is essentially best possible, consider the following graph G given by Nash-Williams (1970). Take the complete bipartite graph with vertex sets $U = \{u_1, \ldots, u_{4k+1}\}$ and $W = \{w_1, \ldots, w_{4k-1}\}$, and add to it the edges u_1u_2, u_3u_4, $u_5u_6, \ldots, u_{4k-1}u_{4k}$ and $u_{4k}u_{4k+1}$. The obtained graph G has $n = 8k$ vertices and minimal degree $2k$. Note that every Hamilton cycle in G has to contain two of the $2k+1$ edges in U, so G has at most $\lfloor (2k+1)/2 \rfloor = k = n/8$ edge-disjoint Hamilton cycles.

Li Hao (1989b) proved a conjecture of Faudree and Schelp that if Ore's condition in Lemma 2.1.1 is satisfied and the graph has *small* minimal degree then there are many edge disjoint cycles.

Theorem 2.2.5. *Let G be a graph with n vertices and minimal degree δ such that $n \geqslant 2\delta^2$ and the degree sum of any two non-adjacent vertices is at least n. Then the graph contains $k = \lfloor (\delta-1)/2 \rfloor$ edge disjoint cycles of lengths $l_1, l_2, \ldots, l_k$, for all $3 \leqslant l_1 \leqslant l_2 \leqslant \cdots \leqslant l_k \leqslant n$.*

2.3. Long cycles

For a graph G, let $C(G)$ be the set of lengths of cycles in G. The *circumference of* G is the length of a longest cycle: $c(G) = \max C(G)$, the *girth* of G is the length of a shortest cycle: $g(G) = \min C(G)$. What do various natural graph parameters (size, minimal degree, connectivity, etc.) tell us about $c(G)$, $g(G)$ and $C(G)$?

Let $x_1 x_2 \cdots x_l$ be a longest path in a graph G, and let $k = \max\{i\colon x_1$ is joined to $x_i\}$. Then $k \geq d(x_1) + 1 \geq \delta(G) + 1$ so, in particular, if $\delta(G) \geq 2$ then $c(G) \geq \delta(G) + 1$. This trivial observation was strengthened considerably by Alon (1986) to a result including Dirac's theorem (Theorem 2.1.2): if $\delta(G) \geq n/k$ then $c(G) \geq \lfloor n/(k-1) \rfloor$. The theorem was extended slightly by Egawa and Miyamoto (1989) and Bollobás and Häggkvist (1990) to the following best possible result.

Theorem 2.3.1. *Suppose $2 \leq k < n$ are integers and G is a graph of order n and minimal degree at least n/k. Then $c(G) \geq n/(k-1)$. Furthermore, for $2 \leq k < n$ there is a graph G of order n such that $\delta(G) = \lceil n/(k-1) \rceil - 1$ and $c(G) = \lceil n/(k-1) \rceil$.*

In fact, recently Bollobás and Brightwell (1993) extended Theorem 2.3.1 to the following result, whose proof turned out to be considerably easier than the proofs of Theorem 2.3.1.

Theorem 2.3.1′. *Let G be a graph of order n with a set W of $w \geq 3$ distinguished vertices. Suppose that every vertex of W has degree at least $d \geq 2$ and let $s = \lceil w/(\lceil n/d \rceil - 1\rceil \geq 3$. Then there is a cycle in G containing at least s vertices of W.*

If we demand that our graph is 2-connected then we can guarantee a considerably longer cycle: as proved by Dirac (1952), if G is 2-connected then $c(G) \geq \min\{|G|, 2\delta(G)\}$. The following extension of a theorem of Pósa (1963) was proved by Bondy (1971a).

Theorem 2.3.2. *Let $3 \leq c \leq n$ and let G be a 2-connected graph of order n with vertex set $\{x_1, x_2, \ldots, x_n\}$ such that $2 \leq d(x_1) \leq d(x_2) \leq \cdots \leq d(x_n)$. Suppose also that if $d_k \leq k < c/2$, $k < l$, $d_l < l$ and $x_k x_l \notin E(G)$ then $k + l \geq c + 1$. Then $c(G) \geq c$.*

Bondy proved also that if in a graph of order n the degree sum of any three independent vertices is at least $m \geq n + 2$ then $c(G) \geq \min\{n, 2m/3\}$, and conjectured the following much stronger result, proved by Fournier and Fraisse (1985) (cf. Theorem 2.1.8.).

Theorem 2.3.3. *Let G be a k-connected graph of order n, where $k \geq 2$, such that the degree sum of any $k + 1$ independent vertices is at least m. Then $c(G) \geq \min\{n, 2m/(k+1)\}$.*

Erdős and Gallai (1959) determined the minimal size of a graph of order n guaranteeing that the circumference is at least c.

Theorem 2.3.4. *Let $3 \leqslant c \leqslant n$. Then the circumference of a graph of order n and size $\lfloor (c-1)(n-1)/2 \rfloor + 1$ at least c.*

A graph G of order n is *pancyclic* if $C(G) = [3, n] = \{3, 4, \ldots, n\}$, i.e., if G contains a cycle of every possible length. We do know that $\lfloor n^2/4 \rfloor$ edges do not guarantee a triangle C_3, and many more edges are needed to guarantee a Hamilton cycle. However, as the following theorem of Bondy (1971b) shows, if a graph has more than $\lfloor n^2/4 \rfloor$ edges then a cycle of length $l > 3$ guarantees a cycle of length $l-1$.

Theorem 2.3.5. *Let G be a graph of order n with more than $\lfloor n^2/4 \rfloor$ edges. Then $c(G) \geqslant \lfloor \frac{1}{2}(n+3) \rfloor$ and $C(G) = [3, c(G)]$. In particular, if G is also Hamiltonian then it is pancyclic.*

How large a minimal degree ensures that a graph G of order n is pancyclic? In view of Theorem 2.3.5 the answer is $\lfloor n/2 \rfloor + 1$, the degree ensuring the existence of a triangle. If G is not bipartite then, as proved by Häggkvist (1982), already $\delta(G) \geqslant (2n+1)/5$ ensures the existence of a triangle. Amar et al. (1983) proved that if G is also Hamiltonian, then the same condition guarantees that the graph is pancyclic, and Shi (1986) showed the following slight extension of this result.

Theorem 2.3.6. *Let G be a non-bipartite Hamiltonian graph of order n such that for any two non-adjacent vertices x and y we have $d(x) + d(y) \geqslant (4n+1)/5$. Then G is pancyclic.*

It is easily seen that Theorem 2.3.6 is best possible. Indeed, let G be the $2k$-regular graph of order $n = 5k$ with vertex set $V = \bigcup_{i=1}^{5} V_i$ where $|V_1| = \cdots = |V_5| = k$ and with edges joining V_i to V_{i+1} for $1 = 1, \ldots, 5$, where $V_6 = V_1$. Then G is not pancyclic because it contains no 4-cycles.

Woodall (1972) determined the minimal number of edges ensuring that a graph G of order n and minimal degree δ satisfies $C(G) \supset [3, l]$. Here we state only a consequence of this result.

Theorem 2.3.7. *Let $3 \leqslant (n+3)/2 \leqslant l \leqslant n$ and let G be a graph of order n and size*

$$\binom{l-1}{2} + \binom{n-l+2}{2} + 1 .$$

Then $C(G) \supset [3, l]$. The bound is best possible.

Although a graph with fewer than $\lfloor n^2/4 \rfloor$ edges cannot be guaranteed to have *any* odd cycles, it *can* be guaranteed to have *even* cycles, both short and long. The following deep and almost best possible result was conjectured by Erdős (1965) and proved by Bondy and Simonovits (1974).

Theorem 2.3.8. *Let k be a natural number. Every graph of order n and size at least $90kn^{1+1/k}$ contains a cycle of length $2l$ for every integer l in the interval $k \leq l \leq kn^{1/k}$.*

2.4. Girth and diameter

What forces a graph to have small girth, i.e., short cycles? Many edges, or almost equivalently, large minimal degree. To study the connection between the minimal degree and the girth, for natural numbers $\delta \leq 2$ and $g \geq 3$ define

$$n(g, \delta) = \min\{|G|: g(G) \geq g \text{ and } \delta(G) \geq \delta\}.$$

A graph of minimal degree δ, girth at least g and order $n(g, \delta)$ is said to be a *(δ, g)-cage*.

It is not entirely immediate that $n(g, \delta) < \infty$, i.e., there are finite graphs of arbitrarily large girth and arbitrarily large minimal degree. However, this does follow from a simple argument using random graphs.

A cycle of length g shows that $n(g, 2) = g$ so we shall assume that $\delta \geq 3$. By estimating the number of vertices at distance d from a vertex or from an edge, one gets the following trivial lower bound on $n(g, \delta)$.

Theorem 2.4.1. *If $\delta \geq 3$ then*

$$n(g, \delta) \geq \begin{cases} 1 + \delta \dfrac{(\delta - 1)^{(g-1)/2} - 1}{\delta - 2} & \text{if } g \text{ is odd}, \\ \dfrac{2(\delta - 1)^{g/2} - 2}{\delta - 2} & \text{if } g \text{ is even}. \end{cases}$$

It is easily seen that in Theorem 2.3.1 equality holds for $\delta = 3$, $g = 3$, 4, 5, 6 and 8, and for $g = 4$ and all $\delta \geq 3$. For example, $n(5, 3) = 10$ is shown by the Petersen graph; the extremal graph for $z(7; 4) = 21$ (see Theorem 1.3.3) shows that $n(6, 3) = 14$ (thus the vertices are the 7 points and 7 lines of the projective plane PG(2, 2), with a point joined to a line if they are incident); the graph $K(\delta, \delta)$ shows that $n(4, \delta) = 2\delta$.

Suppose that $g \geq 3$, $\delta \geq 3$ and G_0 is a graph showing that equality holds in Theorem 2.4.1. If g is odd, say $g = 2D + 1$, then G_0 is δ-regular and has diameter D; also $n(g, \delta)$ is the *maximal* order of a graph with *maximal* degree at most δ and *diameter* at most D. If $g = 2D + 2$ then G_0 is δ-regular and every vertex is within distance D of every edge (in fact, of every pair of vertices); also $n(g, \delta)$ is the *maximal* order of a graph with *maximal* degree at most δ in which *every* vertex is within distance D of *every* edge. Such a graph G_0 is called a *Moore graph of girth g and degree δ.* (If $g = 2D + 1$ then G_0 is also called a Moore graph of diameter D and degree δ.)

There are very few Moore graphs. Results of Hoffman and Singleton (1960), Kárteszi (1960), Feit and Higman (1964), Singleton (1966), Bannai and Ito (1973) and Damerell (1973) show that if there is a Moore graph of girth $g \geqslant 5$ and degree $\delta \geqslant 3$ then either $g = 5$ and $\delta = 3$, 7 or 57, or else $g = 6$, 8 or 12. For $g = 6$ and 8 there is a Moore graph for each finite projective geometry of order δ and dimension 2 and 3.

As there are so few graphs attaining the trivial lower bound in Theorem 2.4.1, what about graphs showing that $n(g, \delta)$ is not much larger than the trivial lower bound. Such graphs are not easy to come by either. The following theorem was proved by Erdős and Sachs (1963) without explicitly constructing a graph showing the inequality.

Theorem 2.4.2. *If $g \geqslant 3$ and $\delta \geqslant 3$ then*

$$n(g, \delta) \leqslant \begin{cases} \dfrac{\delta}{\delta - 2}\{(\delta - 1)^{g-1} - 1\} & \textit{if } g \textit{ is odd}\,, \\ \dfrac{4}{\delta - 2}\{(\delta - 1)^{g-2} - 1\} & \textit{if } g \textit{ is even}\,. \end{cases}$$

Note that for large values of g the upper bound given in Theorem 2.4.2 is about the *square* of the trivial lower bound in Theorem 2.4.1. This huge gap was narrowed by Margulis (1982) by an *explicit construction*: a most welcome success of constructive algebraic methods.

Let $p \geqslant 5$ be a prime and consider $\mathrm{SL}_2(\mathbb{Z}_p)$, the multiplicative group of unimodular 2 by 2 matrices with entries from the field $\mathbb{Z}_p$. Let $A = (\begin{smallmatrix} 1 & 2 \\ 0 & 1 \end{smallmatrix})$ and $B = (\begin{smallmatrix} 1 & 0 \\ 2 & 1 \end{smallmatrix})$ be elements of $\mathrm{SL}_2(\mathbb{Z}_p)$. The *Margulis graph* $M(4, p)$ is the Cayley graph over $\mathrm{SL}_2(\mathbb{Z}_p)$ with respect to the set $\{A, B, A^{-1}, B^{-1}\}$, i.e., $M(4, p)$ has vertex set $\mathrm{SL}_2(\mathbb{Z}_p)$ with a matrix C joined to a matrix D iff $C^{-1}D \in \{A, B, A^{-1}, B^{-1}\}$. Margulis proved that the graph $M(4, p)$ has rather large girth.

Theorem 2.4.3. *Let $\alpha = 1 + \sqrt{2}$, $k \in \mathbb{N}$ and let $p \geqslant 2\alpha^k$ be a prime. Then the graph $M(4, p)$ is a 4-regular graph of order $p(p^2 - 1)$ and girth at least $2k + 1$.*

Note that for large $n = p(p^2 - 1)$ the Margulis graph $M(4, p)$ has girth about $(2/3 \log \alpha) \log n = \log_b n$, where $b = \alpha^{3/2} = 3.751\ldots$, while Theorem 2.4.2 guarantees only a graph of girth about $\log_3 n$.

Margulis (1982) used the same method to construct regular graphs of large girth and arbitrary even degrees. Following Margulis, Imrich (1984) constructed Cayley graphs of factor groups of some subgroups of the modular group to improve the bound in Theorem 2.4.3.

Theorem 2.4.4. *For every $r > 2$ one can effectively construct infinitely many Cayley graphs with n vertices and girth at least*

$$0.4801\ldots(\log n)/\log(d - 1) - 2\,.$$

Furthermore, for $r = 3$ one can have girth at least

$$0.9601\ldots(\log n)/\log 2 - 5\,.$$

It would be of interest to find other explicit constructions for graphs of large girth and large minimal degree.

2.5. *The set of cycles in graphs of given minimal degree*

A graph G of minimal degree $\delta \geq 2$ contains at least $\delta - 1$ cycles of different lengths, i.e., $|C(G)| \geq \delta - 1$. Indeed, let $x_1x_2\cdots x_l$ be a longest path in G and let $x_2, x_{i_1}, x_{i_2}, \ldots, x_{i_k}$ be the neighbours of x_1; then $k \geq \delta - 1$ and for every j, $1 \leq j \leq k$, the graph has a cycle of length i_j, namely $x_1x_2\cdots x_{i_j}$. The graphs $K_{\delta+1}$ and $K(\delta, \delta)$ show that this trivial bound on $|C(G)|$ in terms of δ cannot be improved in general.

However, if $\delta(G) = \delta \geq 3$ and G has large girth then it is easily seen that $|C(G)|$ has to be (much) larger than $\delta - 1$. This suggests that if short cycle lengths are taken with large weights and long cycle lengths are taken with small weights, then the total weight of cycle lengths has to be large if the minimal degree is large. Erdős and Hajnal (see Erdős 1975) proposed taking a cycle of length r with *weight* $1/r$. For a graph G, let

$$S(G) = S(C(G)) = \sum \{1/r\colon\, r \in C(G)\}\,.$$

How large is then

$$f(k) = \inf\{S(G)\colon\, \delta(G) = k\}?$$

The graph $K_{k,k}$, $k \geq 2$ has minimal degree k and its set of cycle lengths is $\{4, 6, \ldots, 2k\}$ so, as $k \to \infty$,

$$f(k) \leq S(K_{k,k}) = \tfrac{1}{2} \sum_{r=2}^{k} \frac{1}{r} = (\tfrac{1}{2} + o(1))\log k\,.$$

Erdős and Hajnal conjectured that $f(k)$ is of order $\log k$. To appreciate the difficulty in proving this conjecture, note that it seems to be difficult to prove that $f(k) \to \infty$ as $k \to \infty$.

This conjecture was proved by Gyárfás et al. (1984).

Theorem 2.5.1. *There are positive constants c and ε such that if $\delta(G) \geq c$ then $S(G) \geq \varepsilon \log \delta(G)$.*

The ingenious and beautiful proof makes good use of the so-called (k, α)-trees. Let T be a rooted tree of height h and levels $L_1, L_2, \ldots, L_h$, where the ith *level* L_i of T is the set of vertices at distance i from the root. This tree T is said to be a (k, α)-*tree* if for $i < h$ every vertex x at level i has at most k neighbours at level

$i+1$ and

α

An impo
assertion.

Theorem 2.5. *bipartite grap* *$4m/7$ even in*

As an imme
another conje

Theorem 2.5.3. *has positive upp*

Proof. By a sim
finite subgraph H
maximal size has

Theorem 2.5.1 ca
degree of G. For

$$h(\alpha) = \text{in}$$

Since every graph G satisfying $e(G) \geqslant \alpha|G|$, i.e., having average de 2α, has a subgraph of minimal degree at least α, Theorem 2.5.1 following result.

Theorem 2.5.1′. *There are positive constants c and ε such that if $\alpha \geqslant c$* $h(\alpha) \geqslant \varepsilon \log \alpha$.

This result gives no information about $h(\alpha)$ for small values of α. Trivially, $h(\alpha) = 0$ for $\alpha \leqslant 1$ but a priori it is not clear that there is no $\alpha_0 > 1$ such that $h(\alpha) = 0$ for $\alpha \leqslant \alpha_0$. Gyárfás et al. (1985) proved that, in fact, $f(\alpha) > 0$ for every $\alpha > 1$.

Theorem 2.5.4. *If k is sufficiently large then* $h(1 + 1/k) \geqslant (300k \log k)^{-1}$.

3. Saturated graphs

A property P of graphs is *monotone increasing* if whenever a graph G has P, so does every graph obtained from G by the addition of some edges. Clearly, if $\mathscr{F}_n$ is the set of minimal graphs of order n having property P then P is determined by the sequence $(\mathcal{F}_n)_{n=1}^{\infty}$, and conversely, if $\mathcal{F}_n$ is a family of graphs for order n then $(\mathcal{F}_n)_{n=1}^{\infty}$ determines a monotone increasing property P: a graph G of order n has P if and only if it contains at least one element of $\mathcal{F}_n$. Using the terminology of the previous section, a graph of order n fails to have property P if it contains no *forbidden subgraph*, i.e., no element of $\mathcal{F}_n$.

A graph G is *P-saturated* or *saturated for P* if G does not have P but any graph obtained from P by the addition of an edge has P. In the first two sections we studied P-saturated graphs with maximal number of edges. Here we shall turn to the lower bound: at least how many edges does a P-saturated graph of order n have? Usually one writes $\operatorname{sat}(n; P)$ for this minimum, i.e., $\operatorname{sat}(n; P) = \min\{e(G)$: $|G| = n$ and G is P-saturated$\}$. Also, the set of extremal graphs is $\operatorname{SAT}(n; P) = \{G$: $|G| = n$, $e(G) = \operatorname{sat}(n; P)$ and G is P-saturated$\}$. If P is given by the sequence $(\mathcal{F}_n)_{n=1}^{\infty}$ then we may write $\operatorname{sat}(n, \mathcal{F}_n)$ and $\operatorname{SAT}(n; \mathcal{F}_n)$ for $\operatorname{sat}(n; P)$ and $\operatorname{SAT}(n; P)$. Also, if $\mathcal{F}_n = \{F_1, \ldots, F_k\}$ then we may write $\operatorname{sat}(n; F_1, \ldots, F_k)$ and $\operatorname{SAT}(n; F_1, \ldots, F_k)$.

3.1. Complete graphs

Erdős et al. (1964) proved the following analogue of Turán's theorem for saturated graphs.

Theorem 3.1.1. *If* $2 \leq r \leq n$ *then* $\operatorname{sat}(n; K_r) = (r-2)(n-1) - \binom{r-2}{2} = (r-2)n - \binom{r-1}{2}$ *and* $\operatorname{SAT}(n; K_r) = \{K_{r-2} + \bar{K}_{n-r+2}\}$, *i.e., the edge set of the unique extremal graph for* $\operatorname{sat}(n; K_r)$ *is the set of all edges incident with a fixed set of* $r-2$ *vertices.*

Proof. Call a graph *K_r-saturated* if it is saturated for the property of containing a K_r subgraph. Furthermore, writing $k_r(G)$ for the number of K_r subgraphs of G, we call G *strongly K_r-saturated* if $k_r(G) < k_r(G^+)$ whenever G^+ is obtained from G by the addition of an edge. Clearly ever K_r-saturated graph is strongly K_r-saturated but a strongly K_r-saturated graph need not be K_r-saturated because it may contain a K_r-subgraph. Note that if G is strongly K_r-saturated then so is every graph obtained from G by the addition of some edges.

The graph $G_n = K_{r-2} + \bar{K}_{n-r+2}$ has $(r-2)n - \binom{r-2}{2}$ edges and it is K_r-saturated. Instead of the claim of the theorem, we shall prove the stronger assertion that every strongly K_r-saturated graph of order n has at least $(r-2)n - \binom{r-1}{2}$ edges, and G_n is the only strongly K_r-saturated graph with n vertices and $(r-2)n - \binom{r-1}{2}$ edges. In fact, as the property of being strongly K_r-saturated is a monotone increasing property, it suffices to prove the latter assertion. We shall do this by induction on $n + r$.

The assertion is trivial if $r = 2$ or $n = r$. Assume then that $3 \leq r < n$ and the result is true for smaller values of $n + r$. Let G be a strongly K_r-saturated graph with n vertices and $(r-2)n - \binom{r-1}{2}$ edges. Let x_1 and x_2 be non-adjacent vertices of G. As G is strongly K_r-saturated, there are vertices $x_3, \ldots, x_r$ such that in the set $\{x_1, x_2, \ldots, x_r\}$ any two vertices are joined to each other, with the exception of x_1 and x_2. Let $H = G/\{x_1, x_2\}$ be the graph obtained from G by identifying x_1

and x_2. Thus $V(H)=\{\bar{x}_2, x_3, \ldots, x_n\}$, for $3 \leq i<j \leq n$ two vertices x_i, x_j are joined in H if and only if they are joined in G, and $\bar{x}_2$ is joined to x_j in H if and only if at least one of x_1 and x_2 joined to x_j in G. Clearly,

$$e(G) \geq e(H)+r-2 .$$

Also, as G is strongly K_r-saturated, so is H. Hence, by the induction hypothesis,

$$e(H) \geq (r-2)(n-1)-\binom{r-1}{2},$$

with equality if and only if $H=G_{n-1}$. Therefore

$$e(G) \geq (r-2)n-\binom{r-1}{2}$$

and if equality holds then $H=G_{n-1}$ and for $i=1$ and 2 the vertices $x_3, \ldots, x_r$ are the only neighbours of x_i in G. It is easily checked that this implies that $G=G_n$, as claimed. □

Let us give another proof of the fact that every strongly K_r-saturated graph of order n has at least $(r-2)n-\binom{r-1}{2}$ edges and so, in particular,

$$\operatorname{sat}(n; K_r) \geq (r-2)n-\binom{r-1}{2}=\binom{n}{2}-\binom{n-r+2}{2}.$$

Let G be a strongly K_r-saturated graph with n vertices. Let $A_1, A_2, \ldots, A_l$ be the (unordered) pairs of vertices not joined to each other. We have to prove that $l \leq \binom{n-r+2}{2}$. For each set A_i there is an r-set $C_i \subset V(G)$ such that $A_i \subset C_i$ and the only two vertices of C_i not joined to each other are the vertices of A_i. Set $B_i=V(G)-C_i$.

Note that $|A_i|=2$, $|B_i|=n-r$ and $A_i \cap B_i=\emptyset$. Furthermore, if $i \neq j$ then $A_i \cap B_i \neq \emptyset$. Indeed, if we had $A_i \cap B_j=\emptyset$ then the set $C_j=V(G)-B_j$ would contain at least two pairs of non-adjacent vertices, namely A_i and A_j. Hence $A_i \cap B_j=\emptyset$ if and only if $i=j$. Thus the required inequality is an immediate consequence of the following theorem of Bollobás (1965).

Theorem 3.1.2. *For two non-negative integers a and b write $w(a, b)=\binom{a+b}{a}^{-1}$. Let $\{(A_i, B_i)\colon i \in I\}$ be a finite collection of finite sets such that $A_i \cap B_j=\emptyset$ if and only if $i=j$. For $i \in I$ set $a_i=|A_i|$ and $b_i=|B_i|$. Then*

$$\sum_{i \in I} w(a_i, b_i) \leq 1$$

with equality if and only if there is a set Y and non-negative integers a and b, such that $|Y|=a+b$ and $\{(A_i, B_i)\colon i \in I\}$ is the collection of all ordered pairs of disjoint subsets of Y with $|A_i|=a$ and $|B_i|=b$ (and so $B_i=Y-A_i$).

In particular, if $a_i=a$ and $b_i=b$ for all $i \in I$ then $|I| \leq \binom{a+b}{a}$. If $a_i=2$ and $b_i=n-r$ for all $i \in I$ then $|I| \leq \binom{n-r+2}{2}$.

Proof. We shall prove the inequality; the case of equality requires a little more work.

We may assume that the sets A_i, B_i are subsets of $[n]$. Call a permutation $\pi = x_1, x_2, \ldots, x_n$ *compatible* with a set-pair (A_i, B_i) if in π every element of A_i precedes every element of B_i. Let N be the number of compatible pairs $(\pi, (A_i, B_i))$. Clearly each set-pair (A_i, B_i) is compatible with

$$\binom{n}{a_i + b_i} a_i!b_i!(n - a_i - b_i)! = n!w(a_i, b_i)$$

permutations π, so

$$N + n! \sum_{i \in I} w(a_i, b_i) .$$

On the other hand, no permutation π is compatible with two set-pairs, say (A_i, B_i) and (A_j, B_j). Indeed, otherwise we may assume that $\max\{k\colon x_k \in A_i\} \leq \max\{k\colon x_k \in A_j\}$. Then $\max\{k\colon x_k \in A_i\} \leq \max\{k\colon x_k \in A_j\} < \min\{k\colon x_k \in B_j\}$ so $A_i \cap B_j = \emptyset$, contradicting our assumption. Hence $N \leq n!$, so

$$N = \sum_{i \in I} w(a_i, b_i)n! \leq n! ,$$

implying the required inequality. □

In fact, Theorem 3.1.2 is an extension of the LYM inequality of Lubell (1966) Yamamoto (1954) and Meshalkin (1963), which, in turn, is an extension of Sperner's (1928) lemma, and the proof given above is just a variant of Lubell's proof of the LYM inequality. To be precise, the LYM inequality is simply the case $B_i = X - A_i$ of Theorem 3.1.2 where X is the ground set.

The original reason for proving Theorem 3.1.2 was to extend Theorem 3.1.1 to hypergraphs: with the appropriate definitions, every k-uniform hypergraph of order n which is saturated for a complete graph with r vertices has at least $\binom{n}{k} - \binom{n-r+k}{k}$ hyperedges.

The proof of Theorem 3.1.2 can be adapted to give us the bipartite version of Theorem 3.1.1, first proved by Bollobás (1967a,b) and Wessel (1966, 1967). An m by n bipartite graph with classes V_1 and V_2 is *strongly saturated for* $K(s, t)$ if the addition of any edge joining V_1 to V_2 creates at least one *new* complete bipartite subgraph with s vertices in V_1 and t vertices in V_2.

Theorem 3.1.3. *Let* $2 \leq s \leq m$ *and* $2 \leq t \leq n$. *An* m *by* n *bipartite graph which is strongly saturated for* $K(s, t)$ *has at least* $mn - (m - s + 1)(n - t + 1)$ *edges. There is only one extremal graph, the* m *by* n *bipartite graph containing all edges joining the two classes except those that join a fixed set of* $n - t + 1$ *vertices in the first class to a fixed set of* $n - t + 1$ *vertices in the second class.*

Duffus and Hanson (1986) studied refinements of the problem of determining sat$(n; K_r)$. Let sat$(n; K_r, \delta)$ be the minimal number of edges in a K_r-saturated graph with n vertices and minimal degree at least δ.

Theorem 3.1.4. *If* $n \geqslant 5$ *then* $\text{sat}(n; K_3, 2) = 2n - 5$ *and if* $n \geqslant 10$ *then* $\text{sat}(n; K_3, 3) = 3n - 15$.

It is easily seen that for $\delta = 2, 3$ the value of sat$(n; K_3, \delta)$ is at most as large as claimed. Given a graph H and a vertex x of H, construct a graph G from H by adding to H a vertex and joining it to the neighbours of x. This graph G is said to have been obtained from H by *duplicating* the vertex x. Note that if H is K_3-saturated then so is G. As the 5-cycle C_5 and the Petersen graph P are K_3-saturated, so are the graphs with n vertices obtained from C_5 and P by repeated duplications of their vertices; these graphs have minimal degrees 2 and 3, and $2n - 5$ and $3n - 15$ edges.

Perhaps for every fixed $\delta \geqslant 1$ one has $\text{sat}(n; K_3, \delta) = \delta n - O(1)$.

3.2. General families

Let us turn to the problem of determining or estimating sat$(n; \mathcal{F})$ for a general family $\mathcal{F}$ of graphs. We know that if no member of $\mathcal{F}$ is bipartite then $\text{ex}(n; \mathcal{F}) \geqslant \lfloor n^2/4 \rfloor$, i.e., there are (maximal) graphs of order n not containing any forbidden graphs which have at least $\lfloor n^2/4 \rfloor$ edges. On the other hand, as the following easy estimate shows, $\text{sat}(n; \mathcal{F}) = O(n)$ for every fixed finite family $\mathcal{F}$.

Theorem 3.2.1. *Let* $\mathcal{F}$ *be a (non-empty) finite family of non-empty graphs and let* $r = \max\{|F|: F \in \mathcal{F}\}$. *Then for* $n \geqslant r$ *we have*

$$\text{sat}(n; \mathcal{F}) \leqslant (r-2)n - \binom{r-1}{2}.$$

Proof. Let us apply induction on r. For $r = 2$ the assertion is trivial because $K_2 \in \mathcal{F}$ so the empty graph $\bar{K}_n$ is $\mathcal{F}$-saturated. Suppose that $r \geqslant 3$ and the result holds for smaller values of r. If $\mathcal{F}$ contains a star $K_{1,s}$, $s \leqslant r = 1$, then a graph containing no member of $\mathcal{F}$ must have maximal degree at most $s - 1$ so

$$\text{sat}(n; \mathcal{F}) \leqslant \frac{s-1}{n} \leqslant \frac{r-2}{2} n \leqslant (r-2)n - \binom{r-1}{2}.$$

Suppose then that no member of $\mathcal{F}$ is a star. Set $\mathcal{F}' = \{F - \{x\}: F \in \mathcal{F}, x \in V(F)\}$. Then $\mathcal{F}'$ is a finite family of non-empty graphs, each with at most $r - 1$ vertices, so by the induction hypothesis,

$$\text{sat}(n-1; \mathcal{F}') \leqslant (r-3)(n-1) - \binom{r-2}{2}.$$

Let H be an extremal graph for sat$(n - 1; \mathcal{F}')$ and let G be obtained from H by adding to it a vertex x and joining x to all $n - 1$ vertices in H. It is trivial that G is

$\mathcal{F}$-saturated, so

$$\mathrm{sat}(n; F) \leq e(G) = \mathrm{sat}(n-1; \mathcal{F}') + n - 1 = (r-2)n - \binom{r-1}{2}. \quad \square$$

Note that for $\mathcal{F} = \{K_r\}$ the simple inequality above is, in fact, an equality. Kászonyi and Tuza (1986) proved the following sharper upper bound for $\mathrm{sat}(n; \mathcal{F})$.

Theorem 3.2.2. *Let $\mathcal{F}$ be a family of non-empty graphs. Set*

$$u = \min\{|U|: F \in \mathcal{F},\ U \subset V(F),\ F - U \text{ is a star}\}$$

and

$$s = \min\{e(F-U): F \in \mathcal{F},\ U \subset V(F),\ F - U \text{ is a star and } |U| = u\}.$$

Furthermore, let p be the minimal number of vertices in a graph $F \in \mathcal{F}$ for which the minimum s is attained. If $n \geq p$ then

$$\mathrm{sat}(n; \mathcal{F}) \leq \left(u + \frac{s-1}{2}\right)n - \frac{u(s+u)}{2}.$$

Proof. We proceed as in proof of Theorem 3.2.1 but this time we apply induction on u. It is again trivial to start the induction: if $u = 0$ then $K_{1,s} \cup \bar{K}_{p-s-1} \in \mathcal{F}$, i.e., $\mathcal{F}$ contains the union of a star with s edges and $p - s - 1$ isolated vertices. Hence, if an $\mathcal{F}$-saturated graph G has an $n \geq p$ vertices than its maximal degree is at most $s - 1$ and so $e(G) \leq (s-1)n/2$. The induction step is as before: the family $\mathcal{F}'$ has parameters $u - 1$, s and $p - 1$ instead of u, s and p, so

$$\begin{aligned}\mathrm{sat}(n; \mathcal{F}) &\leq n - 1 + \left(u - 1 + \frac{s-1}{2}\right)(n-1) - \frac{(n-1)(s+u-1)}{2} \\ &= \left(u + \frac{s-1}{2}\right)n - \frac{u(s+u)}{2}. \quad \square\end{aligned}$$

In the proof above, the star $K_{1,s}$ played a major role; In fact, as pointed out by Kászonyi and Tuza (1986), it is very easy to find the exact value of $\mathrm{sat}(n; K_{1,s})$. Indeed, if G is $K_{1,s}$-saturated, i.e., if G is a maximal graph with maximal degree at most $s - 1$, then any two vertices of degree less than s are joined. This remark and simple calculations imply the exact value of $\mathrm{sat}(n; K_{1,s})$.

Theorem 3.2.3. *If $s + 1 \leq n \leq s + s/2$ then*

$$\mathrm{sat}(n; K_{1,s}) = \binom{s}{2} + \binom{n-s}{2}$$

and if $n > s + s/2$ then

$$\mathrm{sat}(n; K_{1,s}) = \lceil (s-1)n/2 - s^2/8 \rceil.$$

Note that in Theorem 3.2.2 one also has equality for $\mathcal{F} = \{K_r\}$ since then $u = r - s$ and $s = 1$. Furthermore, if $\mathcal{F} = \{C_4\}$, i.e., if the only forbidden graph is the 4-cycle, then $u = 1$ and $s = 2$ so, by Theorem 3.2.2, $\mathrm{sat}(n; C_4) \leqslant 3(n-1)/2$. This bound is also close to being best possible. To obtain a slightly better upper bound, given $n \geqslant 5$, take $t = \lfloor (n-3)/2 \rfloor$ triangles sharing a vertex, and join $p = n - 2t - 1$ new vertices to p vertices of one of these triangles, using independent edges. The obtained graph is C_4-saturated and has n-vertices and $3t + p = \lfloor (3n-5)/2 \rfloor$ edges. As proved by Ollman (1972), this is the best one can do.

Theorem 3.2.4. *If* $n \geqslant 5$ *then* $\mathrm{sat}(n; C_4) = \lfloor (3n-5)/2 \rfloor$.

In conclusion, it is worth remarking that, as noted by Kászonyi and Tuza (1986), the function $\mathrm{sat}(n; \mathcal{F})$ lacks the expected regularity properties. Namely, if $\mathcal{F} \subset \mathcal{F}'$ and $F' \subset F$ then we need not have any of the relations $\mathrm{sat}(n; \mathcal{F}') \leqslant \mathrm{sat}(n; \mathcal{F})$, $\mathrm{sat}(n; F') \leqslant \mathrm{sat}(n; F)$ and $\mathrm{sat}(n; \mathcal{F}) \leqslant \mathrm{sat}(n+1; \mathcal{F})$. Indeed, let F be a K_3 and a K_4 sharing a vertex and let $F' = K_4$. The graph consisting a K_5 and $n - 5$ edges incident with one of the vertices of the K_5 is F-saturated so $\mathrm{sat}(n; F) \leqslant n + 5$. On the other hand, $\mathrm{sat}(n, F') = 2n - 3$ so if $n > 8$ then $\mathrm{sat}(n; F) < \mathrm{sat}(n; F')$. Also, with $\mathcal{F}' = \{F', F\}$ and $\mathcal{F} = \{F\}$ we have $\mathrm{sat}(n; \mathcal{F}) \leqslant n + 5 < \mathrm{sat}(n; \mathcal{F}') = 2n - 3$ for $n > 8$.

3.3. Weakly saturated graphs

Given a family $\mathcal{F}$ of graphs and a graph G, write $k_{\mathcal{F}}(G)$ for the number of subgraphs of G that are isomorphic to members of $\mathcal{F}$. If $\mathcal{F} = \{K_r\}$ then, as before, we write $k_r(G)$ instead of $k_{K_r}(G)$. Call a graph G *weakly* $\mathcal{F}$*-saturated* if there is a sequence of graphs $G_0 = G \subset G_1 \subset \cdots \subset G_m$ such that $V(G_i) = V(G)$, $e(G_i) = e(G_{i-1}) + 1$ and $k_{\mathcal{F}}(G_i) > k_{\mathcal{F}}(G_{i-1})$ for every i, $1 \leqslant i \leqslant m$, and G_m is the complete graph on $V(G)$. Thus G is weakly $\mathcal{F}$-saturated if we can add to it edges one by one in such a way that with each edge we strictly increase the number of $\mathcal{F}$-subgraphs and we stop the process only when our graph is complete. Denote by w-sat$(n; \mathcal{F})$ the minimal number of edges in a weakly $\mathcal{F}$-saturated graph with n vertices.

Since an $\mathcal{F}$-saturated graph is also weakly $\mathcal{F}$-saturated, we have w-sat$(n; \mathcal{F}) \leqslant \mathrm{sat}(n; \mathcal{F})$. As one would expect, w-sat$(n; \mathcal{F})$ can be much smaller than $\mathrm{sat}(n; \mathcal{F})$. For example, let $F = kK_2$, i.e., let F be a set of k independent edges. Tutte's (1947) 1-factor theorem implies easily that every maximal graph with n vertices without k independent edges is of the form $K_s + \bigcup_{i=1}^{q} K_{2n_i+1}$, where $s \geqslant 0$, $n_i \geqslant 0$, $q = n + s - 2k + 2$ and $\sum_{i=1}^{q} (2n_i + 1) = n - s$ (see Bollobás 1978a, Corollary 1.9, p. 58). This implies that if $n \geqslant (5k-2)/2$ then the maximal number of edges in an F-saturated graph with n vertices is

$$\mathrm{ex}(n; kK_2) = \binom{k-1}{2} + (k-1)(n-k+1),$$

as proved by Erdős and Gallai (1961), and if $n \geq 3k - 3$ then the minimal number of edges in an F-saturated graph with n vertices is

$$\mathrm{sat}\{n; kK_2) = 3(k-1)\,,$$

the minimum being given by $k-1$ independent triangles. On the other hand, a graph with $n \geq 2k+1$ vertices and $k-1$ independent edges is weakly kK_2-saturated so

$$\text{w-sat}(n; kK_2) = k - 1\,.$$

Also, it is easily seen that for $n \geq 4$ we have w-sat$(n; C_4) = n$ while Theorem 3.2.4 tells us that sat$(n; C_4) = \lfloor (3n-5)/2 \rfloor$ for $n \geq 5$.

It is fascinating that for $F = K_r$ a weakly F-saturated graph must have at least as many edges as an F-saturated graph: w-sat$(n; K_r) = $ sat$(n; K_r) = (r-2)n - \binom{r-1}{2}$. For very small values of r this is easily seen. For example, a weakly K_3-saturated graph must be connected so w-sat$(n; K_3) \geq n - 1$ and hence w-sat$(n; K_3) =$ sat$(n; K_3) = n - 1$. However, while for sat$(n; K_3)$ there is *just one* extremal graph, the extremal graphs for w-sat$(n; k_3)$ are precisely the *trees*. The large size of the family of extremal graphs even in this trivial case indicates that it is considerably harder to determine w-sat$(n; K_r)$ than sat$(n; K_r)$. This task was accomplished almost twenty years after the original results of Erdős et al. (1964) and Bollobás (1965), by Frankl (1982), Kalai (1984) and Alon (1985).

Theorem 3.3.1. *If* $2 \leq r \leq n$ *then* w-sat$(n; K_r) = (r-2)n - \binom{r-1}{2}$.

To see what is needed to obtain this result, let us return to the proof of Theorem 3.1.1 that led us to Theorem 3.1.2. Let G be a weakly K_r-saturated graph with n vertices and let $G_0 = G \subset G_1 \subset \cdots \subset G_l$ be the sequence showing this. Let A_i be the pair of vertices joined in G_i but not in G_{i-1}. Let C_i be the vertex set of a K_r contained in G_i but not in G_{i-1}, and let $B_i = V(G) - C_i$. Then $|A_i| = 2$ and $|B_i| = n - r$. As $A_i \subset C_i$, we have $A_i \cap B_i$. Furthermore, none of the pairs $A_{i+1}, A_{i+2}, \ldots, A_l$ can be contained in C_i since the vertices in A_i were the last two vertices to be joined in C_i. Hence for $j > i$ we have $A_j \cap B_i \neq \emptyset$. It turns out that these two conditions imply that $l \geq \binom{n-r+2}{2}$ which is the content of Theorem 3.3.1. In fact, Frankl (1982), Kalai (1984) and Alon (1985) proved the appropriate result for all values of $|A_i| = a$ and $|B_i| = b$, which implies the extension of Theorem 3.3.1 for uniform hypergraphs.

Theorem 3.3.2. *Let* $(A_1, B_1), (A_2, B_2), \ldots, (A_l, B_l)$ *be pairs of finite sets such that* $|A_i| = a$, $|B_i| = b$ *and* $A_i \cap B_i = \emptyset$ *for all* i. *Suppose furthermore that* $A_i \cap B_j = \emptyset$ *if* $i > j$. *Then* $l \leq \binom{a+b}{a}$.

The proofs of Theorem 3.3.2, given by Frankl, Kalai and Alon are all rather similar, very beautiful and very unexpected: they make use of exterior powers of

algebras. *With hindsight th*
are clearly dimensions of ex
Lovász (1977) had used exterior
Theorem 3.3.2 is tailor-made for a p
chapter 24 by Frankl.

The following extension of Theorem
Steckin (1982) and proved by Füredi (19

Theorem 3.3.3. *Let* $(A_1, B_1), \dots, (A_l, B_l)$ *b*
$|B_i| \leq b$ *and* $|A_i \cap B_i| \leq c$ *for all i. Suppose*
Then $l \leq \binom{a+b-2c}{a-c}$.

3.4. *Hamilton cycles*

So far we have considered only the function sat(n; $\mathcal{F}$), i.e., we have consi only the case when our forbidden family $\mathcal{F}$ does not depend on n. This sectio devoted to the problem of determining sat(n; $\mathcal{F}_n$) for the prime example of family $\mathcal{F}_n$ depending on n, namely $\mathcal{F}_n = \{C_n\}$. A graph with n vertices is C_n-saturated if it is a maximal non-Hamiltonian graph, i.e., if it is non-Hamiltonian but the addition of any edge creates a Hamilton cycle. The following results were proved by Bondy (1972).

Theorem 3.4.1. *Let G be a maximal non-Hamiltonian graph of order* $n \geq 7$ *with m vertices of degree* 2. *Then G has at least* $(3n + m)/2$ *edges.*

Corollary 3.4.2. *If* $n \geq 7$ *then* sat(n; C_n) $\geq \lceil 3n/2 \rceil$.

When studying sat(n; $\mathcal{F}$) for a fixed family $\mathcal{F}$, it is usually easy to give an upper bound for sat(n; $\mathcal{F}$) and the difficulty lies in proving that the function is at least as large as claimed. Rather curiously, the situation is quite different for sat(n; C_n): the results above are fairly simple, and, as it happens, the lower bound is the actual value of the function, but it is difficult to construct examples showing that sat(n; C_n) is indeed $\lceil 3n/2 \rceil$ if n is not too small.

If n is even then sat(n; C_n) = $\lceil 3n/2 \rceil = 3n/2$ if there is a cubic graph saturated for Hamilton cycles. Since a Hamiltonian cubic graph is 3-edge-colourable, we need a C_n-saturated 4-edge-chromatic cubic graph of order n. In fact, 4-edge-chromatic cubic graphs are not easy to come by: Isaacs (1975) was the first to construct an infinite family of such graphs. By making use of this family, Clark and Entringer (1983) and Clark et al. (1988) proved that sat(n; C_n) = $\lceil 3n/2 \rceil$ for most values of n.

In view of the difficulties with sat(n; C_n), it is unlikely that one could determine even sat(n; C_k) for every pair (k, n). However, getting good bounds on this function may not be hopeless.

Corollary 4.1.4. *Let G_1 and G_2 be graphs with n vertices such that if one has maximal degree $n-1$ then the other has an isolated vertex. If $e(G_1)+e(G_2) \leq 2n-3$ then there is a packing of G_1 and G_2.*

Proof. If the maximal degrees are at most $n-2$ then the result follows from Theorem 4.1.3. Otherwise we may assume that G_1 has a vertex x of degree $n-1$ and G_2 has an isolated vertex y. Placing x on y, there remains to pack $G_1' = G_1 - x$, with $e(G_1) - n + 1$ edges, and $G_2' = G_2 - y$, with $e(G_2)$ edges. Since $e(G_1') + e(G_2') \leq n-2$, it is trivial that there is such a packing, for example, by Theorem 4.1.1. □

This corollary implies immediately the result of Sauer and Spencer mentioned above: if $e(G_1)+e(G_2) \leq 3(n-1)/2$ then there is a packing of G_1 and G_2. Teo (see Yap 1988) extended Theorem 4.1.3 to graphs having a total of $2n-2$ edges. As expected, the number of exceptional pairs increases substantially. For simplicity, we state the result only for $n \geq 13$.

Theorem 4.1.5. *Let G_1 and G_2 be graphs with $n \geq 13$ vertices each such that $\Delta(G_i) \leq n-2$ and $e(G_1)+e(G_2) \leq 2n-2$. For $i = 1, 2, 3$, let H_i be the disjoint union of a star with $n-i-1$ edges and a K_i: $H_i = K_{1,n-i-1} \cup K_i$, let H_4 be a disjoint union of cycles, i.e., a 2-regular graph of order n, and for $n = 3k$ let H_5 be the disjoint union of k triangles: $H_5 = kK_3 = \overline{T_k(n)}$. If $\{G_1, G_2\}$ is not one of the pairs $\{H_1, H_4\}$, $\{H_2, H_4\}$ and $\{H_3, H_3\}$ then there is a packing of G_1 and G_2.*

If one of the graphs to be packed is a tree then one can do considerably better. Extending various earlier results, Slater et al. (1985), proved that if T is a tree of order n, G is a graph of order n and size $n-1$, and neither T nor G is a star then there is a packing of T and G. Furthermore, by making use of Theorem 4.1.3 and this result, Teo and Yap (1987) characterized the graphs of order n and size n which can be packed into the complement of any tree of order n.

It is very likely that, in turn, Theorem 4.1.5 can be extended to graphs with a total of $2n-1$ edges at the expense of a further increase in the set of exceptional pairs but the proof is likely to be forbiddingly cumbersome. However, for the case when the maximal degree is restricted even more, Eldridge (1976) proved the following result. The bound cannot be improved in general.

Theorem 4.1.6. *Let $r \geq 4$ and let G_1 and G_2 be graphs with $n \geq 9r^{3/2}$ vertices and maximal degrees at most $n-r$. If $e(G_1)+e(G_2) < 2n + r(\sqrt{r}-2) - \sqrt{r}$ then there is a packing of G_1 and G_2.*

Rather little is known about packing many graphs with few edges. In particular, if true, the following conjecture of Bollobás and Eldridge (1978) is unlikely to be easy to prove.

Conjecture 4.1.7. For every $k \geq 1$ there is an $n(k)$ such that if $n \geq n(k)$ and G_1,

$G_2, \ldots, G_k$ are graphs with n vertices such that $e(G_i) \leq n-i$ and $\Delta(G_i) \leq n-k$ for every i, $i=1, 2, \ldots, k$, then there is a packing of $G_1, G_2, \ldots, G_k$.

4.2. Graphs of small maximal degree

The results above show that a trivial obstruction to packing is the existence of vertices of very large degrees. If the maximal degrees are known to be small then the existence of a packing follows from much weaker bounds on the total number of edges. Now we shall look for restrictions on the maximal degree only implying the existence of a matching.

The following simple result was announced by Catlin (1974) and proved independently by Sauer and Spencer (1978).

Theorem 4.2.1. *Let G_1 and G_2 be graphs with n vertices such that $\Delta(G_1)\Delta(G_2) < n/2$. Then there is a packing of G_1 and G_2.*

Proof. As for $\Delta(G_1)\Delta(G_2) \leq 1$ there is nothing to prove, we may assume that $\Delta(G_1)\Delta(G_2) \geq 2$ and so $n \geq 5$. Choose an identification of the vertex sets $V(G_1)$ and $V(G_2)$ in which G_1 and G_2 have a minimal number of edges in common. Suppose $V(G_1) = \{x_1, \ldots, x_n\}$, $V(G_2) = \{y_1, \ldots, y_n\}$ and x_i is identified with y_i.

Assume that, contrary to the assertion, G_1 and G_2 share an edge in this identification, say $x_1x_2 \in E(G_1)$ and $y_1y_2 \in E(G_2)$. Let L be the set of indices l such that either $x_2x_j \in E(G_1)$ and $y_jy_l \in E(G_2)$ or else $y_2y_j \in E(G_2)$ and $x_jx_l \in E(G_l)$. Since $x_1x_2 \in E(G_1)$ and $y_1y_2 \in E(G_2)$, we have

$$|L| \leq (\Delta(G_2)-1)\Delta(G_1) + (\Delta(G_1)-1)\Delta(G_2) < n-2\,.$$

Hence there is a natural number k, $3 \leq k \leq n$, such that $k \notin L$. If we flip x_2 and x_k, i.e., if we identify x_2 with y_k and x_k with y_2, then the number of edges common to G_1 and G_2 decreases, contradicting our assumption. □

How far is this result from being best possible? Let $d_1 \leq d_2 < n$ be natural numbers such that $n \leq (d_1+1)(d_2+1)-2$. Let G_1 be a graph such that d_2 of its components are complete graphs of order d_1+1; similarly, let G_2 have d_1 components that are complete graphs of order d_1+1. For example, let $G_1 = d_2K_{d_1+1} \cup K_{d_1-1}$ and $G_2 = d_1K_{d_2+1} \cup K_{d_2-1}$. Note that $\Delta(G_1) = d_1$ and $\Delta(G_2) = d_2$. Suppose that there is a packing of G_1 and G_2. Then every K_{d_1+1} component of G_1 has at least one vertex outside the K_{d_2+1} components of G_2. As there are d_2 components of the form K_{d_1+1} in G_1 but only d_2-1 vertices of G_2 in the K_{d_2+1} components, this is impossible. Hence there is no packing of G_1 and G_2.

Bollobás, Eldridge and Catlin conjectured (see Bollobás 1978b) that the example above is worst possible, i.e., $n/2$ in Theorem 4.2.1 can almost be replaced by n.

Conjecture 4.2.2. Let G_1 and G_2 be graphs with n vertices such that $(\Delta(G_1)+1)(\Delta(G_2)+1) \leq n+1$. Then there is a packing of G_1 and G_2.

At the moment we are very far from a proof of the above conjecture. The following difficult theorem of Hajnal and Szemerédi (1970) provides some evidence for the truth of the conjecture.

Theorem 4.2.3. *Every graph with maximal degree Δ has a $(\Delta+1)$-colouring in which the cardinalities of any two colour classes differ by at most* 1.

Note that the Hajnal–Szemerédi theorem implies Conjecture 4.2.2. in the case when G_2 is of the form $T_r(n)$; in fact, the theorem is more or less equivalent to the conjecture in this case. Indeed, if $G_2 = T_r(n)$ then $\Delta(G_2) = \lceil n/r \rceil - 1$ so if $(\Delta(G_1)+1)(\Delta(G_2)+1) \leq n+1$ then $\Delta(G_2) + 1 = \lceil n/r \rceil \leq (n+1)/(\Delta(G_1)+1)$. Therefore $r \geq \Delta(G_1) + 1$ so Theorem 4.2.3 implies that there is a packing of G_1 and G_2.

One should emphasize that Theorem 4.2.3 itself is a substantial result; various special cases of the theorem had been proved earlier by Dirac (1952), Corrádi and Hajnal (1963), Zelinka (1966), Grünbaum (1968) and Sumner (1969).

Catlin (1977, 1980) proved some special cases of Conjecture 4.2.2, including the following result.

Theorem 4.2.4. *There is a function $f(n) = \mathrm{O}(n^{2/3})$ such that if G_1 and G_2 are graphs with n vertices such that $\Delta(G_1) \leq 2$ and $\Delta(G_2) \leq n/3 - f(n)$, then there is a packing of G_1 and G_2.*

4.3. *Packing trees*

Very little is known about the possibility of packing more than two graphs. The only exception is the case when all the graphs to be packed are trees. In fact A. Gyárfás made the following beautiful conjecture (see Gyárfás and Lehel 1978).

Conjecture 4.3.1. Any sequence of trees $T_2, T_3, \ldots, T_n$ with T_i having i vertices, can be packed into K_n.

Note that the total number of edges of $T_2, T_3, \ldots, T_n$ is $\sum_{i=1}^{n-1} i = \binom{n}{2}$ so in a packing claimed by the conjecture every edge of K_n must belong to precisely one of the trees.

This conjecture, which has come to be known as the tree packing conjecture, is unlikely to be solved in the affirmative in the near future. At the moment the truth of the conjecture is known only in some very special cases. Here we shall give three examples: the first two are due to Gyárfás and Lehel (1978) and the third to Hobbs (1981). Recall that a *star* is a tree of the form $K_{1,m}$, i.e., a tree of diameter 2.

Theorem 4.3.2. *Let $T_2, T_3, \ldots, T_n$ be trees with T_i having i vertices, such that each T_i is a path or a star. Then there is a packing of $T_2, T_3, \ldots, T_n$ into K_n.*

Theorem 4.3.3. *Let $T_2, T_3, \ldots, T_n$ be trees with T_i having i vertices, such that all but at most two of them are stars. Then there is a packing of $T_2, T_3, \ldots, T_n$ into K_n.*

Theorem 4.3.4. *Let $T_2, T_3, \ldots, T_n$ be trees of diameter at most 3 such that T_i has i vertices. Then there is a packing of $T_2, T_3, \ldots, T_n$ into K_n.*

The first two results were extended by Straight (1979). In particular, extending Theorem 4.3.3, he proved the existence of a packing if $\Delta(T_i) \geqslant i-2$ with at most two exceptions. (Note that T_i is a star if $\Delta(T_i) = i-1$.) Furthermore, Straight (1979) verified the tree packing conjecture for $n \leqslant 7$, and Fishburn (1983) proved it for $n \leqslant 9$. Theorem 4.3.4 was also considerably extended by Fishburn (1983). These results indicate that even a disproof of Conjecture 4.3.1 is likely to be difficult.

Packing a family $(T_i)_2^k$ of trees of arbitrary shapes is fairly easy if k is not too large. The following easy result of Bollobás (1983) shows that here we can take $k = \lfloor cn \rfloor$ for some $c > 0$.

Theorem 4.3.5. *Let $(T_i)_2^k$ be a sequence of trees where $k = \lfloor \sqrt{2}n/2 \rfloor$ and T_i has i vertices. Then $T_2, T_3, \ldots, T_k$ can be packed into K_n.*

In fact this result has very little to do with packing, because under the conditions the trees can be packed into K_n one after the other: first we pack T_k, then T_{k-1}, then T_{k-2}, etc.; when we choose a packing of T_i we do not take into account the trees $T_{i-1}, T_{i-2}, \ldots, T_2$. A packing of T_i exists because the graph into which T_i is packed has fairly many edges. In fact, the bound $\lfloor \sqrt{2}n/2 \rfloor$ could be replaced by $\lfloor \sqrt{3}n/2 \rfloor$ if one could prove the following fascinating conjecture proposed by Erdős and Sós in 1963. As it happens, this conjecture was one of the motivations for the conjecture of Gyárfás.

Conjecture 4.3.6. Every graph with n vertices and more than $(k-1)n/2$ edges contains every tree with k edges.

Note that the number of edges is just sufficient to guarantee that the graph contains a path with k edges and a star with k edges.

Rather than strengthen Theorem 4.3.5, perhaps one could prove the following conjecture which is considerably weaker than the tree packing conjecture.

Conjecture 4.3.7. For every $k \geqslant 1$ there is an $n(k)$ such that if $n \geqslant n(k)$ and T_{n-k}, $T_{n-k+1}, \ldots, T_n$ are trees, with T_i having i vertices, then they can be packed into K_n.

4.4. *Packing bipartite graphs*

In this section, we shall prove an attractive result of Hajnal and Szegedy about a special type of packing of bipartite graphs.

Let G_1 and G_2 be n by m bipartite graphs, with bipartitions (U_1, W_1) and $(U_2,$

W_2). We say that there is a *bipartite packing* or simply *packing* of G_1 and G_2 if the n by m complete bipartite graph $K(n, m)$, with bipartition (U, W) contains edge-disjoint subgraphs H_2 and H_2 such that, for $i = 1$, 2, the graph H_i is isomorphic to G_i, with U_i corresponding to U. (Note that, unless $n = m$ *and* G_1 and G_2 are rather sparse, a bipartite packing is just a packing of G_1 and G_2 as *bipartite graphs*, i.e., into $K(n, m)$. This justifies the abbreviated terminology.) Equivalently, G_1 and G_2 have a packing if there are one-to-one maps $f: U_1 \to U_2$ and $g: W_1 \to W_2$ such that if xy is an edge of G_1, with $y \in W_1$, then $f(x)g(y)$ is not an edge of G_2. We shall call the *pair* (f, g) a *packing* of G_1 and G_2.

In the proof of the theorem below, we shall need the following simple consequence of Hall's theorem (see chapter 3) about *matchings* in n by n bipartite graphs.

Lemma 4.4.1. *If the minimal degree of G is at least $n/2$ then G has a matching.*

To keep the notation we need self-explanatory and manageable, for $i = 1, 2$, we denote by $d(U_i)$ the *average* of the degrees of the vertices of G_i belonging to U_i, and by $\Delta(U_i)$ the *maximum* of these degrees. Define $d(W_i)$ and $\Delta(W_i)$ analogously. We are ready to state and prove the promised result of Hajnal and Szegedy (1992).

Theorem 4.4.2. *Suppose that the n by m bipartite graphs G_1, G_2 with bipartition (U_1, W_1), (U_2, W_2), are such that*

$$60 \leqslant \Delta(W_1) < m/20d(U_2)\,,$$

$$60 \leqslant \Delta(W_2) < m/20d(U_1)\,,$$

and, for $i = 1$, 2,

$$\Delta(U_i) \leqslant m/2\log(4m)\,,$$

then there is a bipartite packing of G_1 and G_2.

Proof. Let $f: U_1 \to U_2$ be a one-to-one map. As we shall see in a moment, there is a one-to-one map $g: W_1 \to W_2$ such that (f, g) is a *packing* of G_1 and G_2 if and only if a certain m by n bipartite graph B_f has a matching.

Indeed, define a bipartite graph B_f with bipartition (W_1, W_2) by making y_1y_2 $(y_1 \in W_1,\ y_2 \in W_2)$ an edge of B_f if $g(y_1) = y_2$ does not violate the condition that if $xy \in E(G_1)$ then $f(x)g(y) \notin E(G_2)$. In other words, let $y_1y_2 \in E(B_f)$ if and only if $f(\Gamma(y_1)) \cap \Gamma(y_2) = \emptyset$, i.e., if $y_2 \notin \Gamma(f(\Gamma(y_1)))$, where $\Gamma(x)$ denotes the set of neighbours of a vertex x in the appropriate graph.

In view of Lemma 4.4.1, the theorem follows if we show that for some map f the minimal degree $\delta(B_f)$ of B_f is at least $m/2$. Hence it suffices to show that the *probability* that $\delta(B_f) \geqslant m/2$ for a *random* map f is *strictly positive*. In turn, it suffices to show that the probability, that the degree of a *particular* vertex of B_f is less than $m/2$, is less than $1/2m$. Our aim is then to prove this.

By symmetry it suffices to consider a fixed vertex $y_1 \in W_1$. For simplicity, let $U_2 = [n] = \{1, 2, \ldots, n\}$ and let d_i be the degree of vertex i in G_2. Then

$$d_{B_f}(y_1) = m - |\Gamma(f(\Gamma(y_1)))| \geq m - \sum_{i \in f(\Gamma(y_1))} d_i .$$

Hence, if $d(y_1) = |\Gamma(y_1)| = r$, i.e., y_1 has r neighbours in G_1, then

$$\mathbb{P}\left(f_{B_f}(y_1) < \frac{m}{2}\right) \leq \mathbb{P}_r\left(\sum_{i \in \tau} d_i > \frac{m}{2}\right), \tag{1}$$

where $\mathbb{P}_r$ denotes the probability taken in $[n]^{(r)}$, the space of all r-subsets of $\{1, 2, \ldots, n\}$, and τ is a random element of $[n]^{(r)}$.

With the monotone increasing set system $\mathcal{A} = \{A \in \mathbb{P}(n)\colon \sum_{i \in A} d_i > n/2\}$, inequality (1) becomes

$$\mathbb{P}\left(d_{B_f}(y_1) < \frac{m}{2}\right) \leq \mathbb{P}_r(\mathcal{A}) . \tag{2}$$

Setting $p = 5\Delta(W_1)/4n$, we see that with $q = 1 - p$ we have $pqn \geq 3$ and $r \leq pn - (3pqn)^{1/2}$. A martingale-type inequality implies that, under these conditions,

$$\mathbb{P}_p(\mathcal{A}) \geq \left(1 - \frac{1}{\mathrm{e}}\right)\mathbb{P}_r(\mathcal{A}) \geq \tfrac{1}{2}\mathbb{P}_r(\mathcal{A}), \tag{3}$$

where $\mathbb{P}_p(\mathcal{A})$ is the binomial probability with probability p:

$$\mathbb{P}_p(\mathcal{A}) = \sum_{A \in \mathcal{A}} p^{|A|} q^{n-|A|} .$$

Furthermore, by a standard estimate of the probability in the tail of the binomial distribution,

$$\mathbb{P}_p(\mathcal{A}) < \frac{1}{4m} .$$

Combining this with (1), (2) and (3), we find that

$$\mathbb{P}\left(d_{B_f}(y_1) < \frac{m}{2}\right) < \frac{1}{2m},$$

as desired. □

The conditions in Theorem 4.4.2 are fairly tight: there are many ways of showing this with the aid of random graphs, but we do not go into the details. Note also that in the theorem we proved more than we claimed: for every $f: U_1 \to U_2$ there is a $g: W_1 \to W_2$ such that (f, g) is a bipartite packing.

4.5. The complexity of graph properties

The *complexity* $c(\mathcal{P})$ of a graph property $\mathcal{P}$ is the minimal number of entries in the adjacency matrix of a graph that must be examined in the worst case in order to decide whether the graph has the property or not. It is convenient to spell out this definition in terms of a *game* $\mathcal{P}$ between two players, called the *Constructor and Algy* (or *Hider* and *Seeker*). Denote by $\mathcal{G}^n$ the set of all graphs with a fixed set V of n vertices, say $V = \{1, 2, \ldots, n\}$. Then a *property* $\mathcal{P}$ of graphs on V is a subset of $\mathcal{G}^n$ such that $G \in \mathcal{P}$ whenever a graph isomorphic to G belongs to $\mathcal{P}$. In the game $\mathcal{P}$ Algy asks questions from the Constructor about a graph G on V. Each question is of the form: "Is ab an edge of G?", and each question is answered by the Constructor. When posing a question, Algy takes into account all the information he has received up to that point. The Constructor need not have any particular graph in mind: he may change his choice of graph he is constructing edge by edge according to the questions asked by Algy. The game is over when Algy can decide whether or not the graph the Constructor has been defining will have property $\mathcal{P}$ or not. The aim of the Constructor is to keep Algy guessing for as long as possible. On the other hand, Algy tries to pose as pertinent questions as possible: he would like to decide as soon as possible whether the graph has $\mathcal{P}$ or not. The number of moves of Algy (i.e., the number of questions) in this game, assuming that both players play optimally, is the complexity $c(\mathcal{P})$ of the game $\mathcal{P}$.

Needless to say, the complexity of a digraph property is defined analogously. Moreover, the definition easily carries over to properties of subsets. Given a finite set X, a set system $\mathcal{F}$ on X, i.e., a subset $\mathcal{F}$ of the power set $\mathcal{P}(X)$, is said to be a *property of the subsets of* X. Thus a subset of X has property $\mathcal{F}$ if it belongs to $\mathcal{F}$. Algy's questions are of the form: "Is x an element of our subset $\mathcal{F}$?".

Note that a property of graphs on V is precisely a property of the subsets of $V^{(2)}$, the set of all unordered pairs of elements of V, which is invariant under the permutations (of $V^{(2)}$ induced by the permutations) of V.

A property $\mathcal{F} \subset \mathcal{P}(X)$ is *trivial* if either $\mathcal{F} = 0$ or $\mathcal{F} = \mathcal{P}(X)$; needless to say, one is not interested in trivial properties. As shown by Bollobás and Eldridge (1978), Theorem 4.1.3 concerning the packings of graphs implies a lower bound on the complexity of a non-trivial property of graphs.

Theorem 4.5.1. *The complexity of a non-trivial property of graphs of order n is at least* $2n - 4$.

The bound given in this theorem is unlikely to be best possible although, as the following example due to Best et al. (1974) shows, it does give the correct order of magnitude. A *scorpion graph with n vertices* is a graph containing a path bmt such that b (the *body* vertex) has degree $n-2$, m has degree 2 and t (the *tail* vertex) has degree 1. Note that the graph spanned by the $n-3$ neighbours of b different from m is entirely arbitrary.

Theorem 4.5.2. *The graph property of containing a scorpion graph has complexity at most* $6n$.

For lack of space, in the rest of the section we shall concentrate on elusive properties. A property $\mathcal{F}$ of the subsets of X is *elusive* if $c(\mathcal{F}) = |X|$, i.e., if every element of X must be examined in order to decide whether a subset of X belongs to $\mathcal{F}$ or not. Thus a property $\mathcal{P}$ of graphs of order n is elusive if $c(\mathcal{P}) = \binom{n}{2}$ and a property $\mathcal{Q}$ of digraphs of order n (containing at most one loop at each vertex) is elusive if $c(\mathcal{Q}) = n^2$. Best et al. (1974), Kirkpatrick (1974), Milner and Welsh (1976), Bollobás (1976b) and Yap (1986) have shown that a good many properties of graphs with n vertices are elusive. These properties include the property of being planar (for $n \geqslant 5$), the property of containing a complete graph with r vertices (for $2 < r < n$), the property of having chromatic number k (for $2 \leqslant k \leqslant n$), the property of being 2-connected, the property of being connected and Eulerian, and the property of being connected and containing a vertex of degree 1.

A property $\mathcal{F}$ of the subsets of a set X is *monotone increasing* if $A \in \mathcal{F}$ and $A \subset B \subset X$ imply that $B \in \mathcal{F}$; a *monotone decreasing* property is defined similarly. A property is *monotone* if it is either monotone increasing or monotone decreasing. After some initial difficulties, Aanderaa, Rosenberg, Lipton and Snyder (see Rosenberg 1973 and Lipton and Snyder 1974) advanced the conjecture that every non-trivial monotone property of graphs is close to being elusive in the sense that $c(\mathcal{P}) \geqslant \varepsilon n^2$ for some constant $\varepsilon > 0$. A little later, Best et al. (1974) advanced a sharper form of this conjecture: every non-trivial monotone graph property is elusive. The weaker form of the conjecture was proved by Rivest and Vuillemin (1976).

Theorem 4.5.3. *If $\mathcal{P}$ is a non-trivial property of graphs of order n then $c(\mathcal{P}) \geqslant n^2/16$.*

In fact, Rivest and Vuillemin deduced this result from a theorem claiming that certain set properties are elusive. Given a property $\mathcal{F}$ of subsets of X (i.e., a set system $\mathcal{F} \subset \mathcal{P}(X)$, let $\mathrm{Aut}(\mathcal{F})$ be the group of automorphisms of $\mathcal{F}$, i.e., the group of permutations of X leaving $\mathcal{F}$ invariant: $\mathrm{Aut}(\mathcal{F}) = \{\pi\colon \pi$ is a permutation of X such that if $A \in \mathcal{F}$ then $\pi(A) \in \mathcal{F}\}$.

Theorem 4.5.4. *Let X be a set with p^r elements, where p is a prime, and let $\mathcal{F}$ be a property of subsets of X. If $\mathrm{Aut}(\mathcal{F})$ is transitive on X, $\emptyset \in \mathcal{F}$ and $X \notin \mathcal{F}$ then $\mathcal{F}$ is elusive.*

Encouraged by this beautiful result, Rivest and Vuillemin conjectured that Theorem 4.5.4 was true without any restriction on the number of elements of X. This conjecture has turned out to be false: a counterexample was given by Illies (1978). However, Kahn et al. (1984) proved the exact form of the Best et al. conjecture for prime power values of n.

Theorem 4.5.5. *Let $n = p^r$ where p is a prime. Then every non-trivial monotone property of graphs with n vertices is elusive.*

Kahn et al. used techniques from algebraic topology to prove their beautiful theorem. The crucial step in the proof is that if $\mathcal{F}$ is a non-elusive monotone decreasing property of subsets of X then the abstract simplicial complex of X formed by the elements of $\mathcal{F}$ is collapsible.

The bound $n^2/16$ in Theorem 4.5.3 was improved by Kleitman and Kwiatkowski (1980) to $n^2/9$; Kahn et al. used their theorem to give the even better lower bound $n^2/4 + o(n^2)$.

Let us close with a fascinating conjecture of Kahn et al. (1984) claiming that the analogue of the Best et al. conjecture holds for properties of subsets.

Conjecture 4.5.6. Let $\mathcal{F}$ be a non-trivial monotone property of subsets of X. If $\mathrm{Aut}(\mathcal{F})$ is transitive on X then $\mathcal{F}$ is elusive.

References

Alekseev, V.E.
[1982] Hereditary classes and coding of graphs, *Probl. Cybern.* **39**, 151–164 (in Russian).
[1993] On the entropy values of hereditary classes of graphs, *Discrete Math. Appl.* **3**, 191–199.

Alon, N.
[1985] An extremal problem for sets with applications to graph theory, *J. Combin. Theory A* **40**, 82–89.
[1986] The longest cycle of a graph with a large minimal degree, *J. Graph Theory* **10**, 123–127.

Amar, D., E. Flandrin, I. Fournier and A. Germa
[1983] Hamiltonian pancyclic graphs, *Discrete Math.* **46**, 327.

Babai, L., M. Simonovits and J. Spencer
[1990] Extremal subgraphs of random graphs, *J. Graph Theory* **14**, 599–622.

Baer, R.
[1946] Polarities in finite projective planes, *Bull. Amer. Math. Soc.* **52**, 77–93.

Best, M.R., P. van Emde Boas and H.W. Lenstra
[1974] *A sharpened version of the Aanderaa–Rosenberg Conjecture* (Mathematisch Centrum, Amsterdam).

Bollobás, B.
[1965] On generalized graphs, *Acta Math. Acad. Sci. Hungar.* **16**, 447–452.
[1967a] On a conjecture of Erdős, Hajnal and Moon, *Amer. Math. Monthly* **74**, 178–179.
[1967b] Determination of extremal graphs by using weights, *Wiss. Z. Hochsch. Ilmenau* **13**, 4194, 21.
[1976a] On complete subgraphs of different orders, *Math. Proc. Cambridge Philos. Soc.* **79**, 19–24.
[1976b] Complete subgraphs are elusive, *J. Combin. Theory B* **21**, 1–7.
[1978a] *Extremal Graph Theory, London Mathematical Society Monographs*, Vol. 11 (Academic Press, London) xx + 488pp.
[1978b] Problème, in: *Problèmes Combinatoires et Théorie des Graphes*, eds. J.-C. Bermond, J.-C. Fournier, M. Las Vergnas and D. Sotteau (Editions du CNRS, Paris) p. 437.
[1979] *Graph Theory – An Introductory Course, Graduate Texts in Mathematics*, Vol. 63 (Springer, Berlin) x + 180pp.
[1983] Some remarks on packing trees, *Discrete Math.* **46**, 203–204.
[1985] *Random Graphs* (Academic Press, London) xii + 447pp.
[1986] *Extremal Graph Theory with Emphasis on Probabilistic Methods, CBMS Regional Conference Series in Mathematics*, Vol. 62 (American Mathematical Society, Providence, RI) vii + 64pp.

Bollobás, B., and G.R. Brightwell
[1993] Cycles through specified vertices, *Combinatorica* **13**, 147–155.

Bollobás, B., and S.E. Eldridge
[1978] Packing of graphs and applications to computational complexity, *J. Combin. Theory B* **25**, 105–124.

Bollobás, B., and P. Erdős
[1973] On the structure of edge graphs, *Bull. London Math. Soc.* **5**, 317–321.

Bollobás, B., and R. Häggkvist
[1990] The circumference of a graph with a given minimal degree, in: *A Tribute to Paul Erdős*, eds. A. Baker, B. Bollobás and A. Hajnal (Cambridge University Press, Cambridge) pp. 99–106.

Bollobás, B., and A.G. Thomason
[1994a] Projections of bodies and hereditary properties of hypergraphs, *Bull. London Math. Soc.*, to appear.
[1994b] Hereditary and monotone properties of graphs, to appear.

Bollobás, B., P. Erdős and M. Simonovits
[1976] On the structure of edge graphs II, *J. London Math. Soc.* **12**(2), 219–224.

Bollobás, B., M. Simonovits and E. Szemerédi
[1978] Extremal graphs without large forbidden subgraphs, in: *Advances in Graph Theory*, ed. B. Bollobás (North-Holland, Amsterdam) pp. 29–41.

Bondy, J.A.
[1971a] Large cycles in graphs, *Discrete Math.* **1**, 121–132.
[1971b] Pancyclic graphs I, *J. Combin. Theory B* **11**, 80–84.
[1972] Variations on the hamiltonian theme, *Canad. Math. Bull.* **15**, 57–62.

Bondy, J.A., and V. Chvátal
[1976] A method in graph theory, *Discrete Math.* **15**, 111–135.

Bondy, J.A., and M. Simonovits
[1974] On an upper bound of the graph's chromatic number depending on the graph's degree and density, *J. Combin. Theory B* **23**, 247–250.

Brown, W.G.
[1966] On graphs that do not contain a Thomsen graph, *Canad. Math. Bull.* **9**, 281–285.

Catlin, P.A.
[1974] Subgraphs of graphs, I, *Discrete Math.* **10**, 225–233.
[1977] Embedding subgraphs under extremal degree conditions, in: *Proc. 8th South-Eastern Conf. on Combinatorics, Graph Theory and Optimization, Congress. Numerantium* **XIX**, 139–145.
[1980] On the Hajnal–Szemerédi theorem on disjoint cliques, *Utilitas Math.* **17**, 163–177.

Chung, F.R.K.
[1981] On the decomposition of graphs, *SIAM J. Alg. Discr. Math.* **2**, 1–12.

Chvátal, V.
[1972] On Hamilton's ideals, *J. Combin. Theory B* **12**, 163–168.

Chvátal, V., and P. Erdős
[1972] A note on Hamiltonian circuits, *Discrete Math.* **2**, 111–113.

Chvátal, V., and E. Szemerédi
[1981] On the Erdős–Stone theorem, *J. London Math. Soc.* **23**, 207–214.

Clark, L.H., and R.C. Entringer
[1983] Smallest maximally non-hamiltonian graphs, *Period. Math. Hungar.* **14**, 57–68.

Clark, L.H., R.P. Crane, R.C. Entsinger and H.D. Shapiro
[1986] On smallest maximally nonhamiltonian graphs, in: *Combinatorics, Graph Theory and Computing, Proc. 17th Southeast. Conf., Boca Raton, 1986, Congress. Numerantium* **53**, 215–220.

Corrádi, K., and A. Hajnal
[1963] On the maximal number of independent circuits in a graph, *Acta Math. Acad. Sci. Hungar.* **14**, 423–439.

Dirac, G.A.
[1952] Some theorems on abstract graphs, *Proc. London Math. Soc.* **2**(3), 69–81.
[1963] Extensions of Turán's theorem on graphs, *Acta Math. Sci. Hungar.* **14**, 418–422.

Duffus, D.A., and D. Hanson
[1986] Minimal k-saturated and color critical graphs of prescribed minimum degree, *J. Graph Theory* **10**, 55–67.

Egawa, Y., and T. Miyamoto

[1989] The longest cycles in a graph G with minimum degree at least $|G|/k$, *J. Combin. Theory B* **46**, 356-362.

Eldridge, S.E.

[1976] *Packings of graphs*, Ph.D. Thesis (University of Cambridge, Cambridge).

Erdős, P.

[1962] On the number of complete subgraphs contained in certain graphs, *Publ. Math. Inst. Hungar. Acad. Sci.* **7**, 459–464.

[1965] Extremal problems in graph theory, in: *Theory of Graphs and its Applications*, ed. M. Fiedler (Academic Press, New York) pp. 29–36.

[1967] Extremal problems in graph theory, in: *A Seminar in Graph Theory*, ed. F. Harary (Holt, Rinehart & Winston, New York) pp. 54–64.

[1970] On the graph theorem of Turán (in Hungarian) *Mat. Lapok* **21**, 249–251.

[1975] Some recent progress on extremal problems in graph theory, in: *Proc. Sixth Southeastern Conf. on Combinatorics, Graph Theory and Computing* (Utilitas Math., Winnipeg) pp. 3–14.

Erdős, P., and T. Gallai

[1959] On maximal paths and circuits of graph, *Acta Math. Acad. Sci. Hungar* **10**, 337–356.

[1961] On the minimal number of vertices representing the edges of a graph, *Publ. Math. Inst. Hungar. Acad. Sci.* **6**, 181–203.

Erdős, P., and A. Rényi

[1962] On a problem in the theory of graphs (in Hungarian), *Publ. Math. Inst. Hungar. Acad. Sci.* **7**.

Erdős, P., and H. Sachs

[1963] Reguläre Graphen gegebener Taillenweite mit minimaler Knotenzahl, *Wiss. Z. Martin-Luther Univ. Halle-Wittenberg Math.-Natur. Reihe* **12**, 251.

Erdős, P., and M. Simonovits

[1966] A limit theorem in graph theory, *Studia Sci. Hungar.* **1**, 51–57.

[1982] Compactness results in extremal graph theory, *Combinatorica* **2**, 275–288.

[1983] Supersaturated graphs and hypergraphs, *Combinatorica* **3**, 181–192.

Erdős, P., and A.H. Stone

[1946] On the structure of linear graphs, *Bull. Amer. Math. Soc.* **52**, 1087–1091.

Erdős, P., A. Hajnal and J.W. Moon

[1964] A problem in graph theory, *Amer. Math. Monthly* **71**, 1107–1110.

Erdős, P., A. Rényi and V.T. Sós

[1966a] On a problem of graph theory, *Studia Sci. Math. Hungar.* **1**, 215–235.

Erdős, P., A. Goodman and L. Pósa

[1966b] The representation of graphs by set intersections, *Canad. J. Math.* **18**, 106–112.

Erdős, P., D.J. Kleitman and B.L. Rothschild

[1976] Asymptotic enumeration of K_n-free graphs, in: *Int. Colloq. on Combinatorics, Atti dei Convegni Licei* **17**, Vol. **2**, pp. 19–27.

Erdős, P., P. Frankl and V. Rödl

[1986] The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent, *Graphs and Combinatorics* **2**, 113–121.

Fishburn, P.C.

[1983] Packing graphs with odd and even trees, *J. Graph Theory* **7**, 369–383.

Fournier, I., and P. Fraisse

[1985] On a conjecture of Bondy, *J. Combin. Theory B* **39**, 17–26.

Fraisse, P.

[1986] Circuits including a given set of vertices, *J. Graph Theory* **10**, 553–557.

Frankl, P.

[1982] An extremal problem for two families of sets, *European J. Combin.* **3**, 125–127.

Frankl, P., and J. Pach

[1988] An extremal problem on K_r-free graphs, *J. Graph Theory* **12**, 519–523.

Frankl, P., and V. Rödl

[1984] Hypergraphs do not jump, *Combinatorica* **4**, 149–159.

Frankl, P., and B.S. Stečkin

[1982] Problem 6.41, in: *Combinatorial Analysis, Problems and Exercises* (in Russian) (Nauka, Moscow).

Füredi, Z.

[1983] Graphs without quadrilaterals, *J. Combin. Theory B* **34**, 187–190.

[1984] Geometrical solution of an intersection problem for two hypergraphs, *European J. Combin.* **5**, 133–136.

Grünbaum, B.

[1968] A result in graph-coloring, *Michigan Math. J* **13**, 381–383.

Gyárfás, A., and J. Lehel

[1978] Packing trees of different orders into K_n, *Colloq. Math. Soc. János Bolyai* **18**, 463–469.

Gyárfás, A., J. Komlós and E. Szemerédi

[1984] On the distribution of cycle lengths in graphs, *J. Graph Theory* **8**, 441–462.

Gyárfás, A., H.J. Prömel, E. Szemerédi and B. Voigt

[1985] On the sum of reciprocals of cycle lengths in sparse graphs, *Combinatorica* **5**, 41–52.

Győri, E., and A.V. Kostochka

[1979] On a problem of G.O.H. Katona and T. Tarján, *Acta Math. Acad. Sci. Hungar.* **34**, 321–327.

Häggkvist, R.

[1980] A characterization of non-hamiltonian graphs with large degrees, in: *Combinatorics 79, Part II*, eds. M. Deza, and I.G. Rosenberg, *Ann. Discrete Math.* **9**, 259.

[1982] Odd cycles of specified length in non-bipartite graphs, in: *Advances in Graph Theory*, ed. B. Bollobás, *Ann. Discrete Math.* **13**, 89–100.

[1989] *On the structure of non-hamiltonian graphs I, II*, Manuscript.

[1990] *Divide and conquer I–V*, Manuscripts.

Hajnal, A., and E. Szemerédi

[1970] Proof of a conjecture of Erdős, in: *Combinatorial Theory and its Applications II*, eds. P. Erdős, A. Rényi V.T. Sós, *Colloq. Math. Soc. János Bolyai* **4**, 601–623.

Hajnal, P., and M. Szegedy

[1992] Packing bipartite graphs, *Combinatorica* **12**, 295–302.

Hobbs, A.M.

[1981] Packing trees, in: *Proc. 12th South-Eastern Conf. on Combinatorics, Graph Theory and Computing* (Utilitas Math., Winnipeg) pp. 63–73.

Illies, N.

[1978] A counter-example to the Generalized Aanderaa–Rosenberg Conjecture, *Inform. Process. Lett.* **7**, 154–155.

Imrich, W.

[1984] Explicit construction of regular graphs without small cycles, *Combinatorica* **4**, 53–59.

Isaacs, R.

[1975] Infinite families of nontrivial trivalent graphs which are not Tait colorable, *Amer. Math. Monthly* **82**, 221–239.

Jackson, B.

[1979] Edge-disjoint cycles in regular graphs of large degree, *J. London Math. Soc. (2)* **19**, 13–16.

[1980] Hamilton cycles in regular 2-connected graphs, *J. Combin. Theory B* **29**, 27–46.

Jung, H.A.

[1984] Longest circuits in 3-connected graphs, in: *Finite and Infinite Sets*, eds. A. Hajnal, L. Lovász and V.T. Sós, *Colloq. Math. Soc. János Bolyai* **37**, 403–438.

Kahn, J., M. Saks and D. Sturtevant

[1984] A topological approach to evasiveness, *Combinatorica* **4**, 297–306.

Kalai, G.

[1984] Intersection patterns of convex sets, *Israel J. Math.* **48**, 161–174.

Kászonyi, L., and Zs. Tuza

[1986] Saturated graphs with minimal number of edges, *J. Graph Theory* **10**, 203–210.

Kirkpatrick, D.

[1974] Determining graph properties from matrix representations, in: *Proc. Sixth SIGACT Conf., Seattle*, pp. 84–90.

Kleitman, D., and D.J. Kwiatkowski

[1980] Further results on the Aanderaa Rosenberg conjecture, *J. Combin. Theory Ser B* **28**, 85–95.

Kleitman, D.J., and K.J. Winston

[1980] The asymptotic number of lattices, in: *Combinatorial Mathematics, Optimal Designs and their Applications*, ed. S. Shrivastani, *Ann. Discrete Math.* **6**, 243–249.

Kolaitis, P.H., H.J. Prömel and B.L. Rothschild

[1987] K_{l+1}-free graphs: asymptotic structure and a 0–1 law, *Trans. Amer. Math. Soc.* **303**, 637–671.

Kővári, P., V.T. Sós and P. Turán

[1954] On a problem of K. Zarankiewicz, *Colloq. Math.* **3**, 50–57.

Las Vergnas, M.

[1971] Sur une propriété des arbres maximaux dans un graphe, *C.R. Acad. Sci. Paris Ser A-B* **272**, 1297–1300.

Li, Hao

[1989a] Hamiltonian cycles in regular graphs, *Chinese Sci. Bull.* **34**, 267–268.

[1989b] Edge disjoint cycles in graphs, *J. Graph Theory* **13**, 313–322.

Lipton, R.J., and L. Snyder

[1974] On the Aanderaa–Rosenberg conjecture, *SIGACT News* **6**, 30–31.

Lovász, L.

[1977] Flats in matroids and geometric graphs, in: *Proc. 6th British Combinatorial Conf.*, ed. P.J. Cameron (Academic Press, New York) pp. 45–86.

Lovász, L., and M. Simonovits

[1976] On the number of complete subgraphs of a graph, in: *Proc. 5th British Combinatorial Conf.*, eds. C.St.J.A. Nash-Williams and J. Sheehan (Utilitas Math., Winnipeg) pp. 431–441.

[1983] On the number of complete subgraphs of a graph, II, in: *Studies in Pure Mathematics*, ed. P. Erdős (Birkhäuser, Basel) pp. 459–495.

Lubell, D.

[1966] A short proof of Sperner's lemma, *J. Combin. Theory* **1**, 299.

Mantel, W.

[1907] Problem 28, solution by H. Gouwentak, W. Mantel, J. Teixeira de Mattes, F. Schuh and W.A. Wythoff, *Wiskundige Opgaven* **10**, 60–61.

Margulis, G.A.

[1982] Explicit constructions of graphs without short cycles and low density codes, *Combinatorica* **2**, 71–78.

McGuinness, S.

[1994] The greedy clique decomposition of a graph, *J. Graph Theory* **18**, 427–430.

Meshalkin, L.D.

[1963] A generalization of Sperner's theorem on the number of subsets of a finite set, *Teor. Probab. Ver. Primen.* **8**, 219–220 [1964, *Theory Probab. Appl.* **8**, 204–205].

Milner, E.C., and D.J.A. Welsh

[1976] On the computational complexity of graph theoretical properties, in: *Proc. 5th British Combinatorial Conf.*, eds. C.St.J.A. Nash-Williams and J. Sheehan (Utilitas Math., Winnipeg) pp. 471–487.

Nash-Williams, C.St.J.A.

[1970] Hamilton lines in graphs whose vertices have sufficiently large valencies, in: *Combinatorial Theory and Its Applications*, eds. P. Erdős, A. Rényi and V.T. Sós (North-Holland, Amsterdam) pp. 813–819.

[1971] Edge-disjoint Hamiltonian circuits in graphs with vertices of large valency, in: *Studies in Pure Mathematics*, eds. R. Rado and L. Mirsky (Academic Press, London) pp. 157–183.

Nordhaus, E.A., and B.M. Stewart

[1963] Triangles in an ordinary graph, *Canad. J. Math.* **15**, 33–41.

Ollman, L.T.

[1972] $K_{2,2}$-saturated graphs with a minimal number of edges, in: *Proc. 3rd South-Eastern Conf. on Combinatorics, Graph Theory and Computing* (Utilitas Math., Winnipeg) pp. 367–392.

Ore, O.

[1960] Note on Hamiltonian circuits, *Amer. Math. Monthly* **67**, 55.

Pósa, L.

[1963] On the circuits of finite graphs, *Publ. Math. Inst. Hungar. Acad. Sci.* **8**, 355–361.

Prömel, H.J., and A. Steger

[1991] Excluding induced subgraphs: quadrilaterals, *Random Structures and Algorithms* **2**, 55–71.

[1992] Excluding induced subgraphs III: a general asymptotic, *Random Structures and Algorithms* **3**, 19–31.

[1993a] Excluding induced subgraphs II: Extremal graphs, *Discrete Appl. Math.* **44**, 283–294.

[1993b] The assymptotic structure of *H*-free graphs, in: *Graph Structure Theory, Contemporary Mathematics*, Vol. 147, eds. N. Robertson and P. Seymour (AMS, Providence, RI) pp. 167–178.

Pyber, L.

[1986] Clique coverings of graphs, *Combinatorica* **6**, 393–401.

Reiman, I.

[1958] Über ein Problem von K. Zarankiewicz, *Acta Math. Acad. Sci. Hungar.* **9**, 269–279.

Rivest, R.L., and J. Vuillemin

[1975] A generalization and proof of the Aanderaa–Rosenberg conjecture, in: *Proc. 7th Annu. ACM Symp. on Theory of Computing, Albuquerque, NM* (Assoc. Comput. Mach., New York) pp. 6–11.

Rosenberg, A.G.

[1973] On the time required to recognize properties of graphs: a problem, *SIGACT News* **5**, 15–16.

Sauer, N.

[1971] A generalization of a theorem of Turán, *J. Combin. Theory B* **10/11**, 109–112.

Sauer, N., and J. Spencer

[1978] Edge disjoint placement of graphs, *J. Combin. Theory B* **25**, 295–302.

Shi, Ronghua

[1986] *Zentralblatt* **566**, 30–31.

Simonovits, M.

[1968] A method for solving extremal problems in graph theory, in: *Theory of Graphs*, eds. P. Erdős and T. Gallai (Academic Press, New York) pp. 279–310.

[1974] Extremal graph problems with symmetrical extremal graphs. Additional chromatic conditions, *Dicc. Math.* **7**, 349–376.

Skupien, Z.

[1979] On maximal non-Hamiltonian graphs, *Rostock Math. Kolloq.* **11**, 97–106.

Slater, P.J., S.K. Teo and H.P. Yap

[1985] Packing a tree with a graph of the same size, *J. Graph Theory* 9, 213–216.

Sperner, E.

[1928] Ein Satz über Untermengen einer endlichen Menge, *Math. Z.* **27**, 544–548.

Straight, H.J.

[1979] Packing trees of different sizes into a complete graph, in: *Topics in Graph Theory*, ed. F. Harary, *Ann. New York Acad. Sci.* **328**, 190–192.

Sumner, D.P.

[1969] On a problem of Erdős, in: *Recent Progress in Combinatorics*, ed. W.T. Tutte (Academic Press, New York) pp. 319–322.

Szemerédi, E.

[1975] On a set containing no *k* elements in arithmetic progression, *Acta Arithm.* **27**, 199–245.

[1978] Regular partitions of graphs, in: *Problèmes Combinatoires et Théorie des Graphes*, eds. J.-C. Bermond, J.-C. Fournier, M. Las Vergnas and D. Sotteau (Editions du CNRS, Paris) pp. 399–401.

Teo, S.K.

[1985] *Packing of graphs*, M.Sc. Thesis (Department of Mathematics, National University of Singapore).

Teo, S.K., and H.P. Yap

[1987] Two theorems on packing of graphs, *European J. Combin.* **8**, 199–207.

Turán, P.

[1941] On an extremal problem in graph theory (in Hungarian), *Mat. Fiz. Lapok* **48**, 436-452.

[1954] On the theory of graphs, *Colloq. Math.* **3**, 19 30.

Tutte, W.T.

[1947] The factorisation of linear graphs, *J. London Math. Soc.* **22**, 107 111.

Wessel, W.

[1966] Über eine Klasse paarer Graphen, I: Beweis einer Vermutung von Erdős, Hajnal and Moon, *Wiss. Z. Techn. Hochsch. Ilmenau* **12**, 253-256.

[1967] Über eine Klasse paarer Graphen, II: Bestimmung der Minimalgraphen, *Wiss. Z. Techn. Hochsch. Ilmenau* **13**, 423-426.

Woodall, D.R.

[1972] Sufficient conditions for circuits in graphs, *Proc. London Math. Soc.* **24**(3), 739-755.

Yamamoto, K.

[1954] Logarithmic order of free distributive lattices, *J. Math. Soc. Jpn.* **6**, 343-353.

Yap, H.P.

[1986] *Some Topics in Graph Theory, London Mathematical Society Lecture Note Series*, Vol. 108 (Cambridge University Press, Cambridge) vi + 230pp.

[1988] Packing graphs – a survey, in: *Proc. First Japan Conf. on Graph Theory and Applications, Hakone, 1986, Discrete Math.* **72**, 395-404.

Zarankiewicz, K.

[1951] Problem P 101, *Colloq. Math.* **2**, 116-131.

Zelinka, B.

[1966] On the number of independent complete subgraphs, *Publ. Math. Debrecen* **13**, 95-97.

CHAPTER 24

Extremal Set Systems

P. FRANKL

CNRS, Paris, France

Contents

HANDBOOK OF COMBINATORICS
Edited by R. Graham, M. Grötschel and L. Lovász

1. Introduction

Let X be an n-element set and $\mathcal{F} \subset 2^X$ a family of *distinct* subsets of X. Suppose that the members of $\mathcal{F}$ satisfy some given conditions. What is the maximum (minimum) value of $|\mathcal{F}|$? This is the generic problem in extremal set theory and we shall try to give an overview of the existing results and methods. Here is the simplest result:

Theorem 1.1. *If $F \cap F' \neq \emptyset$ holds for all F, $F' \in \mathcal{F} \subset 2^X$, then $|\mathcal{F}| \leqslant 2^{n-1}$.*

Proof. For each $A \subset X$ either A or $X \setminus A$ (or both) are absent from $\mathcal{F}$. Thus $|\mathcal{F}| \leqslant \frac{1}{2}2^n = 2^{n-1}$. □

2. Basic definitions and conventions

For s, t positive integers, $s \geqslant 2$, a family $\mathcal{F}$ is called *s-wise t-intersecting* if $|F_1 \cap \cdots \cap F_s| \geqslant t$ holds for all $F_1, \ldots, F_s \in \mathcal{F}$. If $t = 1$, then t is omitted. Also if $s = 2$, then s-wise is omitted. Thus, "intersecting" means 2-wise 1-intersecting.

A family $\mathcal{F}$ is called *k-uniform* or a k-graph if $|F| = k$ for all $F \in \mathcal{F}$.

The *size* of a family $\mathcal{F}$ is $|\mathcal{F}|$ and it is often denoted simply by m. The members of $\mathcal{F}$ are also called *edges*. Let $\binom{X}{k}$ denote the family of all k-element subsets of X.

For $\mathcal{F} \subseteq 2^X$, set $\mathcal{F}^{(i)} = \{F \in \mathcal{F}\colon |F| = i\}$, and $f_i = |\mathcal{F}^{(i)}|$. In this case $f = (f_0, \ldots, f_n)$ is called the *f-vector* of $\mathcal{F}$.

Let $[n]$ denote $\{1, \ldots, n\}$, $[i, j] = \{l\colon i \leqslant l \leqslant j\}$. Usually we suppose $X = [n]$.

For $i \in X$, define $\mathcal{F}(i) = \{F \setminus \{i\}\colon i \in F \in \mathcal{F}\}$, the *link* of i; $\mathcal{F}(\bar{i}) = \{F \in \mathcal{F}\colon i \notin F\}$.

The *degree* $d_{\mathcal{F}}(i)$ is simply $|\mathcal{F}(i)|$; $\delta(\mathcal{F})$ and $\Delta(\mathcal{F})$ denote the minimum and maximum degree, respectively.

$\mathcal{F}^c = \{X \setminus F\colon F \in \mathcal{F}\}$ is the *complementary family* of $\mathcal{F}$.

$\mathcal{F} \subset 2^X$ is called *hereditary* if $E \subset F \in \mathcal{F}$ implies $E \in \mathcal{F}$. (Note that $\emptyset \in \mathcal{F}$.)

$\mathcal{F} \subset 2^X$ is called a *filter* if $\mathcal{F}^c$ is hereditary.

The lth *shadow* $\sigma_l(\mathcal{F})$ of a family $\mathcal{F}$ is defined by:

$$\sigma_l(\mathcal{F}) = \left\{G \in \binom{X}{l}\colon \exists F \in \mathcal{F}, G \subset F\right\}.$$

$\partial(\mathcal{F}) = \{G \subseteq X\colon G \notin \mathcal{F}; \exists F \in \mathcal{F}, |G \Delta F| = 1\}$ is called the *boundary* of $\mathcal{F}$.

$\nu(\mathcal{F})$, the *matching number* of $\mathcal{F}$, is the maximum number of pairwise disjoint edges in $\mathcal{F}$; $\nu(\mathcal{F}) = \infty$ if $\emptyset \in \mathcal{F}$.

$\tau(\mathcal{F})$, the *covering number* of $\mathcal{F}$, is the minimum cardinality of a set T with $T \cap F \neq \emptyset$ for all $F \in \mathcal{F}$; $\tau(\mathcal{F}) = \infty$ if $\emptyset \in \mathcal{F}$.

$\mathcal{F}$ is called *ν-critical* if $\nu(\mathcal{G}) > \nu(\mathcal{F})$ holds for every family obtained from $\mathcal{F}$ by replacing one of its edges by a proper subset of it.

$\mathcal{F}$ is called *τ-critical* if $\tau(\mathcal{G}) < \tau(\mathcal{F})$ for all $\mathcal{G} \subset \mathcal{F}$.

$\mathcal{F}$ is called an *antichain* if $F \not\subseteq F'$ holds for all F, $F' \in \mathcal{F}$.

Define the *reverse lexicographic order* $<_L$ on 2^X by $A <_L B$ if $A \subset B$ or $\max\{x \in A \setminus B\} < \max\{x \in B \setminus A\}$.

Let $\mathcal{L}(m,k)$ ($\mathcal{R}(m,k)$) be the largest (smallest) m members of $\binom{[n]}{k}$ in the reverse lexicographic order.

Note that $\mathcal{R}(\binom{s}{k}, k) = \binom{[s]}{k}$.

We call $\mathcal{F}$, $\mathcal{G}$ *cross-intersecting* if $F \in \mathcal{F}$ and $G \in \mathcal{G}$ implies $F \cap G \neq \emptyset$.

$\mathcal{F}$ is called a *sunflower* of size m and with *center* C if $F \cap F' = C$ for all distinct F, $F' \in \mathcal{F}$ and $|\mathcal{F}| = m$.

$\mathcal{F}$ is said to be *intersection-closed* if F, $F' \in \mathcal{F}$ implies $F \cap F' \in \mathcal{F}$.

We close this section with a conjecture of Frankl (1979).

Conjecture 2.1. If $\mathcal{F}$ is intersection-closed, $|\mathcal{F}| \geq 2$, then $\delta(\mathcal{F}) \leq |\mathcal{F}|/2$ holds.

3. Basic theorems

The oldest result in extremal set theory is Sperner's Theorem.

Theorem 3.1 (Sperner 1928). *If $\mathcal{F} \subset 2^X$ is an antichain, then $|\mathcal{F}| \leq \binom{n}{\lfloor n/2 \rfloor}$ with equality if and only if $\mathcal{F} = \binom{X}{\lfloor n/2 \rfloor}$ or $\mathcal{F} = \binom{X}{\lceil n/2 \rceil}$ holds.*

Recent research on antichains belongs to the theory of partially ordered sets. We refer the reader to chapter 8 or the book by Engel and Gronau (1985).

The maximum size of intersecting k-graphs was determined in 1938 by Erdős, Ko and Rado although they did not publish their result until much later.

Theorem 3.2 (Erdős et al. 1961). *If $\mathcal{F} \subset \binom{X}{k}$ is t-intersecting, $k > t \geq 1$, $n \geq n_0(k,t)$, then $|\mathcal{F}| \leq \binom{n-t}{k-t}$.*

From the work of Frankl (1978) and Wilson (1984) we know that the conclusion holds if and only if $n \geq (k-t+1)(t+1)$.

Another classical result is due to Erdős and Rado (1960).

Theorem 3.3. *If $\mathcal{F} \subset \binom{X}{k}$, $|\mathcal{F}| > k!(r-1)^k$, then $\mathcal{F}$ contains a sunflower of size r.*

Erdős (1981) offers \$1000 for a proof that the same holds for $|\mathcal{F}| > c(r)^k$, where $c(r)$ is an appropriate constant.

Probably the single most important result in finite set theory is the Kruskal–Katona Theorem, which was proved by Kruskal (1963) and Katona (1966) [see also Lindström (1967), where a somewhat weaker statement is proved].

Theorem 3.4. *If $\mathcal{F} \subset \binom{X}{k}$ is a family of size m, then for all $l < k$, $|\sigma_l(\mathcal{F})| \geq \sigma_l(\mathcal{R}(m,k))$.*

Evaluating $|\sigma_l(\mathscr{R}(m,k))|$ one can get explicit bounds, which, however, are unsuitable for computations. The irregular behaviour of the Kruskal–Katona function is explained in Frankl et al. (1995c). Lovász (1979) gives the following weaker but more convenient version.

Theorem 3.5. *Let $\mathscr{F}\subset\binom{X}{k}$, $|\mathscr{F}|=m$, and define the real number $x\geqslant k$ by $m=\binom{x}{k}$. Then $|\sigma_l(\mathscr{F})|\geqslant\binom{x}{l}$ holds for all $l<k$.*

A simple common proof of Theorems 3.4 and 3.5 was given by Frankl (1984). The values of m and k for which $\mathscr{R}(m,k)$ is the unique optimal family in Theorem 3.4 were determined independently by Füredi and Griggs (1986) and Mörs (1985).

Hilton (1976) noticed that the Kruskal–Katona Theorem can be restated in the following form.

Theorem 3.6. *If $\mathscr{F}\subset\binom{X}{k}$ and $\mathscr{G}\subset\binom{X}{l}$ are cross-intersecting, then so are $\mathscr{L}(|\mathscr{F}|,k)$ and $\mathscr{L}(|\mathscr{G}|,l)$.*

Theorem 3.7 (Matsumoto and Tokushige 1989). *If $\mathscr{F}\subset\binom{X}{k}$ and $\mathscr{G}\subset\binom{X}{l}$ are cross-intersecting and $n\geqslant 2k\geqslant 2l$, then $|\mathscr{F}||\mathscr{G}|\leqslant\binom{n-1}{k-1}\binom{n-1}{l-1}$.*

Another important theorem on shadows is due to Katona (1964).

Theorem 3.8. *If $\mathscr{F}\subset\binom{X}{k}$ is t-intersecting, then for all $k-t\leqslant l\leqslant k$ one has*

$$|\sigma_l(\mathscr{F})|/|\mathscr{F}|\geqslant\binom{2k-t}{l}\Big/\binom{2k-t}{k}\geqslant 1\,.$$

Katona used this theorem to determine the maximum size of t-intersecting families $\mathscr{F}\subset 2^X$, which we will discuss in section 5. Katona showed also that the case $t=1$ of the Erdős–Ko–Rado Theorem 3.2 is an easy consequence of Theorem 3.8.

The *discrete isoperimetric problem* can be stated as follows: given m, determine $\min\{|\partial\mathscr{F}|\colon \mathscr{F}\subset 2^X,\ |\mathscr{F}|=m\}$.

A *ball* with *center* A and *radius* r is the family $\mathscr{B}(A,r)=\{B\subseteq X\colon |A\,\Delta B|\leqslant r\}$.

If $\mathscr{B}(A,r)\subseteq\mathscr{F}\subseteq\mathscr{B}(A,r+1)$, then $\mathscr{F}$ is called a *generalized ball.* Harper (1966) shows that generalized balls have minimum boundary.

Theorem 3.9. *For every $\mathscr{F}\subset 2^X$ there exists a generalized ball $\mathscr{G}\subset 2^X$ of the same size with $|\partial(\mathscr{F})|\geqslant|\partial(\mathscr{G})|$.*

A short proof of this result was given by Frankl and Füredi (1981).

For $\mathscr{F}\subset\binom{X}{k}$ one defines its k-boundary $\kappa(\mathscr{F})$ by:

$$\kappa(\mathscr{F})=\left\{G\in\binom{X}{k}\colon\ G\not\in\mathscr{F};\ \exists F\in\mathscr{F},\ |G\,\Delta F|=2\right\}.$$

One of the outstanding open problems is the isoperimetric problem for $\binom{X}{k}$.

Open Problem 3.10. Given m, determine $\min\{|\kappa(\mathcal{F})|: \mathcal{F} \subset \binom{X}{k}, |\mathcal{F}| = m\}$.

The next result is due to Kleitman (1966a).

Theorem 3.11. *Let $\mathcal{C}$, $\mathcal{D} \subset 2^X$ be hereditary. Then*

$$|\mathcal{C} \cap \mathcal{D}| \geq |\mathcal{C}|\,|\mathcal{D}|/2^n .$$

Proof. Apply induction on n, the case $n = 0$ being trivial. Set $c_0 = |\mathcal{C}(\bar{n})|$, $c_1 = |\mathcal{C}(n)|$, $d_0 = |\mathcal{D}(\bar{n})$, and $d_1 = |\mathcal{D}(n)|$. Then

$$\begin{aligned} |\mathcal{C} \cap \mathcal{D}| &= |\mathcal{C}(n) \cap \mathcal{D}(n)| + |\mathcal{C}(\bar{n}) \cap \mathcal{D}(\bar{n})| \\ &\geq (c_1 d_1 + c_0 d_0)/2^{n-1} \quad \text{(by induction)} \\ &= (c_0 + c_1)(d_0 + d_1)/2^n + (c_0 - c_1)(d_0 - d_1)/2^n . \end{aligned}$$

Using $\mathcal{C}(n) \subseteq \mathcal{C}(\bar{n})$ and $\mathcal{D}(n) \subseteq \mathcal{D}(\bar{n})$, $(c_0 - c_1)(d_0 - d_1) \geq 0$, which completes the proof. □

By now there are many generalizations of Theorem 3.11, some of which are discussed in chapter 8.

4. Basic tools

The most useful tool for investigating s-wise t-intersecting families is an operation called shifting, which was introduced by Erdős et al. (1961).

Definition 4.1. For $\mathcal{F} \subset 2^X$ and $1 \leq i < j \leq n$, define the (i, j)-*shift* S_{ij} by $S_{ij}(\mathcal{F}) = \{S_{ij}(F): F \in \mathcal{F}\}$, where

$$S_{ij}(F) = \begin{cases} (F\setminus\{j\}) \cup \{i\} =: \tilde{F} & \text{if } j \in F, i \not\in F \text{ and } \tilde{F} \not\in \mathcal{F} , \\ F & \text{otherwise} . \end{cases}$$

Some of the useful properties of the (i, j)-shift are summarized by the next lemma.

Lemma 4.2.
(i) $|\mathcal{F}| = |S_{ij}(\mathcal{F})|$ *and* $|F| = |S_{ij}(F)|$;
(ii) $\sigma_l(S_{ij})(\mathcal{F})) \subseteq S_{ij}(\sigma_l(\mathcal{F}))$;
(iii) *if $\mathcal{F}$ is s-wise t-intersecting, then so is $S_{ij}(F)$*;
(iv) $\nu(S_{ij}(\mathcal{F})) \leq \nu(\mathcal{F})$.

Iterating the (i, j)-shift for all $1 \leq i < j \leq n$ will eventually produce a family $\mathcal{G}$ which is invariant with respect to the (i, j)-shift.

Definition 4.3. We call $\mathcal{G}$ *stable* if $S_{ij}(\mathcal{G}) = \mathcal{G}$ for all $1 \leq i < j \leq n$. The followi result is straightforward to show.

Proposition 4.4. *$\mathcal{G}$ is stable if and only if for all $G \in \mathcal{G}$, $1 \leq i < j \leq n$, with $j \in ($ $i \not\in G$, $(G\backslash\{j\} \cup \{i\})$ is also in $\mathcal{G}$.*

A variation of the (i, j)-shift, called down-shift was defined by Kleitma (1966).

Definition 4.5. For $\mathcal{G} \subset 2^X$ and $i \in X$, define the *down-shift* D_i by $D_i(\mathcal{G})$ $\{D_i(G)\colon G \in \mathcal{G}\}$, where

$$D_i(G) = \begin{cases} G - \{i\} & \text{if } i \in G \in \mathcal{G} \text{ and } (G - \{i\}) \not\in \mathcal{G}\,, \\ G & \text{otherwise}\,. \end{cases}$$

Define the *trace* $\mathcal{F}|_Y = \{F \cap Y\colon F \in \mathcal{F}\}$.

Some important properties of the down-shift are summarized in the ne lemma; property (ii) is due to Kleitman (1966), and (iii) to Frankl (1983).

Lemma 4.6.

(i) $|D_i(\mathcal{G})| = |\mathcal{G}|$;

(ii) *if $|F \,\Delta F'| \leq d$ holds for all F, $F' \in \mathcal{F}$, then the same holds for $D_i(\mathcal{F})$*;

(iii) $|D_i(\mathcal{G})|_Y| \leq |\mathcal{G}|_Y|$ *for all $i \in X$ and $Y \subset X$.*

Iterating the down-shift again produces an invariant family.

Proposition 4.7. *$D_i(\mathcal{G}) = \mathcal{G}$ holds for all $i \in X$ if and only if $\mathcal{G}$ is hereditary.*

Let us use this proposition to give a simple proof of the following result whic was discovered independently by three sets of authors: Sauer; Shelah and Perles and Vapnik and Chervonenkis.

Theorem 4.8. *If $|\mathcal{F}| > \sum_{0<i<r} \binom{n}{i}$, then there is some $R \in \binom{X}{r}$ with $\mathcal{F}|_R = 2^R$.*

Proof. Suppose that $|\mathcal{F}|_R| < 2^r$ for all $R \in \binom{X}{r}$. In view of Lemma 4.6 (iii) we ma apply the down-shift to $\mathcal{F}$, and by Proposition 4.7 obtain a complex $\mathcal{G}$, stil satisfying $|\mathcal{G}|_R| < 2^r$ for all $R \in \binom{X}{r}$. However, since $\mathcal{G}$ is hereditary, this implie $|G| < r$ for all $G \in \mathcal{G}$, whence $|\mathcal{G}| \leq \sum_{0 \leq i < r} \binom{n}{i}$ follows. □

We point out that the largest r such that there exists a set $R \in \binom{x}{r}$ with $\mathcal{F}|_R = 2^r$ is called the Vapnik–Chervonenkis dimension of $\mathcal{F}$. This concept has foun interesting applications in combinatorial and computational geometry, an learnability theory (e.g., see Blumer et al. 1989, Clarkson et al. 1988, and Linia et al. 1991).

Another important tool for investigating families of finite sets is the inclusion matrices.

Definition 4.9. For $\mathcal{F} \subset 2^X$, the $|\sigma_j(\mathcal{F})|$ by $|\mathcal{F}|$ matrix $M(j, \mathcal{F})$ has its rows indexed by $G \in \sigma_j(\mathcal{F})$ and its columns by $F \in \mathcal{F}$, and its general entry is

$$m(G, F) = \begin{cases} 1 & \text{if } G \subseteq F\,, \\ 0 & \text{if } G \not\subseteq F\,. \end{cases}$$

Simple computation gives the next result.

Proposition 4.10. (i) *$M(j, \mathcal{F})^T M(j, \mathcal{F})$ is an $|\mathcal{F}|$ by $|\mathcal{F}|$ matrix with general entry*

$$n(F, F') = \binom{|F \cap F'|}{j}\,;$$

(ii) *$M(j, \mathcal{F})^T M(j, \mathcal{F}^c)$ is an $|\mathcal{F}|$ by $|\mathcal{F}|$ matrix with general entry*

$$n(F, F') = \binom{|F \backslash F'|}{j}\,.$$

Definition 4.11. $\mathcal{F} \subset \binom{X}{k}$ is called *k-partite* if there exists a partition $X = X_1 \cup \cdots \cup X_k$ with $|F \cap X_i| = 1$ for all $F \in \mathcal{F}$, $1 \leqslant i \leqslant k$.

A simple but useful result of Erdős and Kleitman (1968) is the following lemma.

Lemma 4.12. *Every k-graph $\mathcal{F}$ contains a k-partite k-graph $\mathcal{G}$ with $|\mathcal{G}|/|\mathcal{F}| \geqslant k!/k^k$.*

Definition 4.13. For a k-partite $\mathcal{F} \subset \binom{X}{k}$ and $F \in \mathcal{F}$, define $\Pi(F, \mathcal{F}) = \{\Pi(F \cap F')\colon F \neq F' \in \mathcal{F}\}$, where $\Pi(A) = \{i\colon A \cap X_i \neq \emptyset\}$. (Thus $\Pi(F, \mathcal{F}) \subset 2^{[k]}$.)

Definition 4.14. We call $\mathcal{F} \subset 2^X$ *r-complete* if for all distinct F, $F' \in \mathcal{F}$ there is a sunflower of size r and with center $F \cap F'$ formed by members of $\mathcal{F}$.

Füredi (1983) discovered the following lemma which has since proved very useful.

Lemma 4.15. *There exists a positive constant $c = c(k, l)$ such that every $\mathcal{F} \subset \binom{X}{k}$ has a k-partite subfamily F^* satisfying*

(i) $|\mathcal{F}^*| \geqslant c|\mathcal{F}|$;

(ii) *$\mathcal{F}^*$ is k-partite with $\Pi(\mathcal{F}^*) = \Pi(F, \mathcal{F}^*)$ being the same for all $F \in \mathcal{F}^*$;*

(iii) *$\mathcal{F}^*$ is l-complete.*

Proposition 4.16 (Deza). *If $l > k$ in Lemma 4.15, then $\Pi(\mathcal{F}^*)$ is intersection-closed.*

Proof. Take D', $D'' \in \Pi(\mathcal{F}^*)$, and choose F, F', $F'' \in \mathcal{F}^*$ with $D' = \Pi(F \cap F')$, $D'' = \Pi(F \cap F'')$. Let $G_1, \ldots, G_{k+1}$ and $H_1, \ldots, H_{k+1}$ be members of $\mathcal{F}^*$ forming sunflowers with centers $C' = F \cap F'$ and $C'' = F \cap F''$, respectively. The sets $G_1 \backslash C'$, $G_2 \backslash C', \ldots, G_{k+1} \backslash C'$ are pairwise disjoint; thus one of them, say $G_1 \backslash C'$, is disjoint from C''. Similarly, $H_1 \backslash C'', \ldots, H_{k+1} \backslash C''$ are pairwise disjoint, implying that one of them, say $H_1 \backslash C''$, is disjoint from G_1. Now $G_1 \cap H_1 = C' \cap C''$, implying $\Pi(C' \cap C'') = D' \cap D'' \in \Pi(\mathcal{F}^*)$ (in the last step we used that $\mathcal{F}^*$ is k-partite). □

Having some information about $\Pi(\mathcal{F}^*)$, one can often use it to get upper bounds on $|\mathcal{F}^*|$ (and thus for $|\mathcal{F}|$).

Proposition 4.17.

$$|\mathcal{F}^*| \leq \binom{n}{\tau(\Pi(\mathcal{F}^*)^c)} .$$

Proof. Let $T \subset [1,k]$ be a minimal set with $T \cap ([1,k] \backslash P) \neq \emptyset$ for all $P \in \Pi(\mathcal{F}^*)$. That is, $|T| = \tau(\Pi(\mathcal{F}^*)^c)$ and $T \not\subseteq P$ for all $P \in \Pi(\mathcal{F}^*)$. For each $F \in \mathcal{F}^*$, let $T(F)$ be the unique subset of F with $\Pi(T(F)) = T$. Since $T \not\subseteq \Pi(F \cap F')$ for distinct F, $F' \in \mathcal{F}^*$, all the $T(F)$ are distinct subsets of X, which concludes the proof. □

5. Intersecting families

Let us define the family $\mathcal{K}(n,t)$ as follows:

$$\mathcal{K}(n,t) = \begin{cases} \{K \subseteq X \colon |K| \geq (n+t)/2\} & \text{if } n+t \text{ is even,} \\ \{K \subset X \colon |K \cap [2,n]| \geq ((n-1)+t)/2\} & \text{if } n+t \text{ is odd.} \end{cases}$$

It is easy to check that $\mathcal{K}(n,t)$ is t-intersecting. Let us state and prove Katona's Theorem.

Theorem 5.1 (Katona 1964). *If $\mathcal{K} \subset 2^X$ is t-intersecting, then $|\mathcal{K}| \leq |\mathcal{K}(n,t)|$, and moreover, for $t \geq 2$, equality holds only if $\mathcal{K}$ is (isomorphic to) $\mathcal{K}$ (n,t).*

Proof. Let us start with a definition. $\mathcal{F} \subset 2^X$ has the *t-union property* if $|F \cup F'| \leq n-t$ for all $F, F' \in \mathcal{F}$.

Now $\mathcal{F} = \mathcal{K}^c$ has the t-union property.

We shall deal only with the case $n-t$ odd; the even case is slightly easier. Set $s = (n+1-t)/2$. Recall that f_i is the number of i-sets in $\mathcal{F}$.

Claim 5.2.

$$f_i + \frac{i+t-1}{i} f_{n-i-t+1} \leq \binom{n}{i}, \quad 0 \leq i \leq s .$$

Proof. Let us consider $\sigma_i(\mathcal{K}^{(i+t-1)})$. If A is in this family, then $A \not\subseteq \mathcal{F}^{(i)}$ since otherwise $|A \cup B^c| = n - t + 1$ holds for $B \in \mathcal{K}^{(i+t-1)}$, $A \subset B$, violating the hypothesis. Thus, $f_i + |\sigma_i(\mathcal{K}^{(i+t-1)})| \leq \binom{n}{i}$. Since $\mathcal{K}$ is t-intersecting we may apply Theorem 3.8 to get

$$|\sigma_i(\mathcal{K}^{(i+t-1)})| \geq f_{n-(i+t-1)}(i + t - 1)/i ,$$

which yields Claim 5.2. □

Proof of Theorem 5.1 (continued). For $i = s$ one has $n - i - t + 1 = i$ and from Claim 5.2, $f_s \leq \binom{n-1}{s-1}$ follows. Adding up this inequality, together with Claim 5.2 applied to $0 \leq i < s$ and noting $f_i = 0$ for $i > n - t$, we obtain

$$|\mathcal{F}| \leq \binom{n-1}{s-1} + \sum_{0 \leq i < s} \binom{n}{i} = 2 \sum_{0 \leq i < s} \binom{n-1}{i} = |\mathcal{K}(n, t)| .$$

If $t \geq 2$, then $(i + t - 1)/i > 1$; thus in the case of equality $\mathcal{K}^{(i+t-1)} = \emptyset$ and consequently $\mathcal{F}^{(i)} = \binom{[n]}{i}$ for $i < s$, which gives already the bulk of the proof of uniqueness. To conclude the proof one notes that $\mathcal{F}^{(s)}$ is intersecting, and $f_s = \binom{n-1}{s-1}$, so by the uniqueness part of the Erdős–Ko–Rado Theorem (which we will discuss subsequently) $\mathcal{F}^{(s)} = \{F \in \binom{[n]}{s}: 1 \in F\}$. This implies $\mathcal{F} = \mathcal{K}(n, t)^c$. □

Theorem 5.3 (Kleitman 1966b). *Suppose that $\mathcal{F} \subset 2^X$ satisfies $|F \Delta F'| \leq n - t$ for all F, $F' \in \mathcal{F}$. Then $|\mathcal{F}| \leq |\mathcal{K}(n, t)|$.*

Proof. In view of Lemma 4.6 we may repeatedly replace $\mathcal{F}$ by $D_i(\mathcal{F})$. Thus by Proposition 4.7 we may suppose that $\mathcal{F}$ is hereditary. Since for arbitrary G, $G' \in \mathcal{F}$ we can take subsets F, $F' \in \mathcal{F}$ with $F \Delta F' = G \cup G'$, $\mathcal{F}$ has the t-union property. Thus Theorem 5.3 follows from Theorem 5.1. □

Let us define some intersecting families $\mathcal{H}(k, s)$ for $2 \leq s \leq k$:

$$\mathcal{H}(k, s) = \left\{H \in \binom{[n]}{k}: 1 \in H \text{ and } [2, s + 1] \cap H \neq \emptyset\right\}$$
$$\cup \left\{H \in \binom{[n]}{k}: [2, s + 1] \subseteq H\right\} .$$

It is easy to check that for $n > 2k$, $|\mathcal{H}(k, 3)| < \cdots < |\mathcal{H}(k, k)|$ holds.

Checking the degrees one sees that

$$\Delta(\mathcal{H}(k, s)) = \binom{n-1}{k-1} - \binom{n-1-s}{k-1} .$$

Theorem 5.4 (Frankl 1987a). *Let $\mathcal{F} \subset \binom{[n]}{k}$ be intersecting, $n > 2k$. If*

$$\Delta(\mathcal{F}) \leq \binom{n-1}{k-1} - \binom{n-1-s}{k-1}$$

holds for some $2 \leq s \leq k$, then $|\mathcal{F}| \leq |\mathcal{H}(k,s)|$. Moreover, equality holds only if $\mathcal{F}$ is isomorphic to $\mathcal{H}(k,s)$, or $s = 3$ and $\mathcal{F}$ is isomorphic to $\mathcal{H}(k,2)$.

Let $\mathcal{F} \subset \binom{X}{k}$ be an intersecting family in which the intersection of all sets satisfies $\bigcap \mathcal{F} = \emptyset$. That is, for each $i \in X$ there is some $F \in \mathcal{F}$ with $i \not\in F$. This implies

$$d_{\mathcal{F}}(i) \leq \binom{n-1}{k-1} - \binom{n-k-1}{k-1}.$$

Thus Theorem 5.4 implies:

Theorem 5.5 (Hilton and Milner 1967). *If $\mathcal{F} \subset \binom{X}{k}$ is an intersecting family with $\bigcap \mathcal{F} = \emptyset$, then for $n > 2k$, $|\mathcal{F}| \leq |\mathcal{H}(n,k)|$ with equality holding if and only if $\mathcal{F} \cong \mathcal{H}(n,k)$, or $k = 3$ and $\mathcal{F} \cong \mathcal{H}(n,2)$.*

Let us mention that the restriction $n > 2k$ is essential because for $n = 2k$ a family $\mathcal{F} \subset \binom{[2k]}{k}$ is intersecting if and only if it contains no set together with its complement. Thus there are $2^{\binom{2k-1}{k-1}}$ distinct intersecting families with $\binom{2k-1}{k-1}$ members in $\binom{[2k]}{k}$. Can they be regular, i.e., $d_{\mathcal{F}}(i) = d$ for some d and all $i \in [2k]$? Simple computation shows that $d = \frac{1}{2}\binom{2k-1}{k-1}$ which is an integer if and only if k is not a power of 2.

Theorem 5.6 (Brace and Daykin 1972). *There exists a regular intersecting family of maximum size $\binom{2k-1}{k-1}$ in $\binom{[2k]}{k}$ if and only if k is not a power of 2.*

Definition 5.7. Let A denote the set of all even integers $2k$ such that there exists an intersecting family $\mathcal{F} \subset \binom{[2k]}{k}$ with $|\mathcal{F}| = \binom{2k-1}{k-1}$ and such that the automorphism group $\mathrm{Aut}(\mathcal{F})$ is transitive on $[2k]$.

Theorem 5.8 (Cameron et al. 1989).

(i) *If $a \in A$ then $ab \in A$ for $b \in A$ and for b odd.*

(ii) *$4a + 2 \in A$ for all positive integers a.*

(iii) *$3 \cdot 2^k \not\in A$ for $k \geq 2$.*

Actually, an even number $2k \in A$ if and only if there is a transitive permutation group on $[2k]$ in which every 2-element has a fixed point.

Conjecture 5.9. $a \cdot 2^d \not\in A$ holds for every fixed a and $d > d_0(a)$.

The maximum size of t-intersecting families in $\binom{X}{k}$ is determined by the

Erdős–Ko–Rado Theorem for $n \geqslant n_0(k, t)$. However, for $t \geqslant 2$ this leaves open a whole range of cases $2k - t < n < (k - t + 1)(t + 1)$. Define the t-intersecting families $\mathcal{A}_i = \mathcal{A}_i(n, k, t)$ for $0 \leqslant i \leqslant k - t$ by:

$$\mathcal{A}_i = \left\{ A \in \binom{[n]}{k} : |A \cap [2i + t]| \geqslant i + t \right\} .$$

Conjecture 5.10 (Frankl 1978). If $\mathcal{F} \subset \binom{X}{k}$ is t-intersecting and $n \geqslant 2k - t$, $k \geqslant t \geqslant 2$, then

$$|\mathcal{F}| \leqslant \max_i |\mathcal{A}_i| .$$

Let us prove a weaker statement.

Proposition 5.11. *If $\mathcal{F} \subset \binom{X}{k}$ is t-intersecting and $n \geqslant 2k - t$, then*

$$|\mathcal{F}| \leqslant \binom{n}{k - t} .$$

Proof. In view of Lemma 4.2 we may assume that $\mathcal{F}$ is stable. The following lemma is often useful.

Lemma 5.12 (Frankl 1978). *If $\mathcal{F} \subset \binom{X}{k}$ is t-intersecting and stable, then $|F \cap F' \cap [2k - t]| \geqslant t$, i.e., $\mathcal{F}_{[2k-t]}$ is t-intersecting.*

Proof. Suppose that Lemma 5.12 is not true and choose a counterexample (F, F') with $|F \cap [2k - t]|$ as large as possible. Fix $j \in F \cap F'$ with $j > 2k - t$. If $i \notin F \cup F'$ for some $i \in [2k - t]$, then replacing (by Proposition 4.4) F by $(F \backslash \{j\}) \cup \{i\}$ contradicts the maximality of $|F \cap [2k - t]|$. Thus $F \cup F' \supseteq [2k - t]$. However,

$$|(F \cup F') \cap [2k - t]| \leqslant |F| + |F'| - |F \cap F' \cap [2k - t]| < 2k - t ,$$

a contradiction. □

Proof of Proposition 5.11 (continued). Apply induction on k. The case $k = t$ is trivial. Also, in the case $n = 2k - t$ one has $|\mathcal{F}| \leqslant \binom{2k-t}{k} = \binom{2k-t}{k-t}$. Let $n > 2k - t$ and define

$$\mathcal{F}_i = \left\{ A \in \binom{[2k - t]}{i} : \exists F \in \mathcal{F}, A = F \cap [2k - t] \right\} .$$

Then by Lemma 5.12 and induction,

$$|\mathcal{F}_i| \leqslant \binom{2k - t}{i - t}$$

holds. This implies

$$|\mathcal{F}| \leqslant \sum_{i=t}^{k} \binom{2k - t}{i - t} \binom{n - 2k + t}{k - i} = \binom{n}{k - t} .$$ □

Theorem 5.13 (Kleitman 1966a). *Let* $\mathcal{F}_1, \ldots, \mathcal{F}_r \subset 2^X$ *be intersecting. Then* $|\mathcal{F}_1 \cup \cdots \cup \mathcal{F}_r| \leqslant 2^n - 2^{n-r}$.

Proof. Apply induction on r; the case $r = 1$ is just Theorem 1.1. We can assume that $\mathcal{F}_1, \ldots, \mathcal{F}_r$ are filters. Consider $\mathcal{F} = \mathcal{F}_1 \cup \cdots \cup \mathcal{F}_{r-1}$. By the induction hypothesis, $|\mathcal{F}| \leqslant 2^n - 2^{n-r+1}$. Also $|\mathcal{F}_r| \leqslant 2^{n-1}$ by Theorem 1.1. Since $\mathcal{F}$ and $\mathcal{F}_r$ are both filters, using Theorem 3.11 we obtain $|\mathcal{F} \cap \mathcal{F}_r| \geqslant |\mathcal{F}| \cdot |\mathcal{F}_r|/2^n$. Summarizing,

$$|\mathcal{F}_1 \cup \cdots \cup \mathcal{F}_r| = |\mathcal{F}_r| + |\mathcal{F}| - |\mathcal{F} \cap \mathcal{F}_r| \leqslant |\mathcal{F}_r| + |\mathcal{F}| - \frac{|\mathcal{F}_r||\mathcal{F}|}{2^n} .$$

The right-hand side is monotone increasing in both $|\mathcal{F}|$ and $|\mathcal{F}_r|$. Thus we get an upper bound by substituting $|\mathcal{F}_r| = 2^{n-1}$ and $|\mathcal{F}| = 2^n - 2^{n-r+1}$. This completes the proof. □

Another application of Theorem 3.11 is the following result which was proved originally in a different way by Daykin and Lovász, and Schönheim.

Theorem 5.14. *If* $\mathcal{F} \subset 2^X$ *is intersecting and has the "union property"* $(F \cup F' \neq X$ *for* $F, F' \in \mathcal{F})$, *then* $|\mathcal{F}| \leqslant 2^{n-2}$.

Proof. Define

$$\mathcal{F}^* = \{G \subseteq X\colon \exists F \in \mathcal{F}, F \subseteq G\} , \quad \text{and} \quad \mathcal{F}_* = \{G\colon \exists F \in \mathcal{F}, G \subseteq F\} .$$

Then $\mathcal{F}^*$ is an intersecting filter and $\mathcal{F}_*$ is hereditary and has the union property. Using Theorems 1.1 and 3.11 we deduce

$$|\mathcal{F}| \leqslant |\mathcal{F}^* \cap \mathcal{F}_*| \leqslant |\mathcal{F}^*||\mathcal{F}_*|/2^n \leqslant 2^{n-2}$$ □

It was shown by Frankl (1975) (proving a conjecture of Katona) that the maximum size of an intersecting family having the t-union property is $|\mathcal{K}(n-1, t)|$.

Example. Let $t, t' \geqslant 1$ and suppose $X = Y \cup Y'$ is a partition with $|Y| \geqslant t$, $|Y'| \geqslant t'$. Let $\mathcal{A} \subset 2^Y$ be a copy of $\mathcal{K}(|Y|, t)$ and $\mathcal{B} \subset 2^{Y'}$ be a copy of $\mathcal{K}(|Y'|, t')^c$. Set $\mathcal{C}(Y) = \{H \subset X\colon H \cap Y \in \mathcal{A}, H \cap Y' \in \mathcal{B}\}$. Then $\mathcal{C}$ is t-intersecting and has the t'-union property.

Conjecture 5.15. If $\mathcal{F} \subset 2^X$ is t-intersecting and has the t'-union property, then $|\mathcal{F}| \leq |\mathcal{C}(Y)|$ for an appropriate $Y \subset X$.

This conjecture can be found in Frankl's dissertation of 1976 and first appeared in English in Bang et al. (1981).

Let us close this section with the following important conjecture of Chvàtal.

Conjecture 5.16. If $\mathscr{C}$ is hereditary, $\mathscr{F} \subset \mathscr{C}$, and $\mathscr{F}$ is intersecting, then $|\mathscr{F}| \leq \Delta(\mathscr{C})$.

For some partial results and references on this conjecture see Miklòs (1984).

6. Families with prescribed intersection sizes

Let $L = \{l_0, \ldots, l_{s-1}\} \subseteq [0, k-1]$ with $l_0 < l_1 < \cdots < l_{s-1}$.

Definition 6.1. A family $\mathscr{F} \subseteq \binom{X}{k}$ is called an (n, k, L)-*system*, or an L-*system* for short, if $|F \cap F'| \in L$ holds for all distinct $F, F' \in \mathscr{F}$. For example, a t-intersecting family is an L-system with $L = \{t, t+1, \ldots, k-1\}$.

Definition 6.2. Let $m(n, k, L)$ denote the maximum size of an (n, k, L)-system.

The next fundamental theorem was first proved by Deza.

Theorem 6.3 (Deza et al. 1978).

$$m(n, k, L) \leq \prod_{l \in L} (n-l)/(k-l) \quad \textit{for } n > n_0(k, L).$$

We remark that an L-system $\mathscr{F} \subset \binom{X}{k}$ with $L = \{0, 1, \ldots, t-1\}$, is called a partial t-design and clearly Theorem 6.3 holds for all $n \geq k$ in this case. A celebrated result of Rödl (1985) is the following.

Theorem 6.4.

$$m(n, k, \{0, 1, \ldots, t-1\}) = (1 - \mathrm{o}(1))\binom{n}{t} \Big/ \binom{k}{t},$$

where $k \geq t > 0$ are fixed and $n \to \infty$.

Taking $L = \{t, t+1, \ldots, k-1\}$, one sees that for $n > n_0(k, L)$, Theorem 6.3 extends Theorem 3.2.

Definition 6.5. We say that Theorem 6.3 is *asymptotically exact* (respectively, gives the *correct exponent*) if

$$\limsup_{n \to \infty} m(n, k, L) \Big/ \prod_{l \in L} \frac{n-l}{k-l}$$

is equal to one (respectively, is positive). For example, Theorem 6.4 shows that Theorem 6.3 is asymptotically exact for all $k \geq t > 0$ and $L = \{0, 1, \ldots, t-1\}$.

Definition 6.6. $L - a = \{l - a \colon l \in L\}$.

In view of the following result of Deza et al. (1978) we may suppose in what follows that $0 \in L$.

Proposition 6.7. *$m(n, k, L) = m(n - l_0, k - l_0, L - l_0)$ for $n > n_0(k, L)$.*

The next result gives some values of k and L for which Theorem 6.3 is asymptotically exact.

Theorem 6.8 (Frankl and Rödl 1985). *Let $d \geq t > 0$ and let q be a prime power. Then Theorem 6.3 is asymptotically exact for*

$$k = q^d\,, \qquad L = \{0, 1, \ldots, q^{t-1}\}$$

and

$$k = (q^d - 1)/(q - 1)\,, \qquad L = \{(q^i - 1)/q - 1)\colon i = 0, 1, \ldots, t - 1\}\,.$$

Definition 6.9. $a(k, L) = \sup\{\alpha\colon \limsup_{n\to\infty} m(n, k, L)n^{-\alpha} > 0\}$.

That is, $a(k, L) \leq |L|$ with equality if and only if Theorem 6.3 gives the correct exponent. Clearly $a(k, L) \geq 1$ for all $\emptyset \neq L \subseteq [0, k - 1]$.

Conjecture 6.10. There exist positive constants $c(k, L)$ and $\bar{c}(k, L)$ for all k, L such that

$$c(k, L)n^{a(k,L)} < m(n, k, L) < \bar{c}(k, L)n^{a(k,L)}\,.$$

Theorem 6.11 (Frankl 1986b). *For every rational number $\alpha \geq 1$ there are infinitely many choices of k and L for which Conjecture 6.10 holds with $a(k, L) = \alpha$.*

One can use Lemma 4.15 and Proposition 4.16 to get upper bounds on $a(k, L)$. Let $\mathcal{F}$ be an (n, k, L)-system and apply Lemma 4.15 with $l = k + 1$ to get the intersection-closed family $\mathcal{A} = \prod(\mathcal{F}^*) \subseteq 2^{[k]}$.

We call a set $B \subset [k]$ a *base* (for $\mathcal{A}$) if $B \not\subseteq A$ for all $A \in \mathcal{A}$ but no proper subset of B has this property. Also, $b(\mathcal{A}) = \min\{|B|\colon B \text{ is a base}\}$.

For $D \subseteq [k]$ define $\langle D\rangle = \bigcap\{A\colon D \subseteq A \in (\mathcal{A} \cup \{[k]\})\}$. That is, $\langle D\rangle = [k]$ if and only if D contains some base for $\mathcal{A}$.

Since $\mathcal{F}^*$ is an L-system, $|A| \in L$ for all $A \in \mathcal{A}$. By Proposition 4.16, there is at most one l_0-element set in $\mathcal{A}$ and one can prove easily that $b(\mathcal{A}) \leq |L|$. In fact, more is true. For elementary properties of matroids, we refer the reader to chapter 9.

Theorem 6.12 (Frankl 1982). *$b(\mathcal{A}) \leq |L| - 1$ unless $\mathcal{A} \cup [k]$ forms the flats of a matroid of rank $|L|$. In this case $b(\mathcal{A}) = |L|$.*

Proof. We apply induction on k; the case $k = 1$ is trivial. Suppose that $b(\mathcal{A}) =$

$|L|$. Define $\mathscr{A}_i = \{A \in \mathscr{A}: |A| = l_i\}$, $0 \leq i < s = |L|$. We have to show that for every $A \in \mathscr{A}_i$ and $x \in [k] \setminus A$, there is a unique member of $\mathscr{A}_{i+1}$ containing both x and A. Define $\tilde{A} = \bigcap \{A': (A \cup \{x\}) \subseteq A' \in \mathscr{A}\}$. Then $\tilde{A} \in \mathscr{A}$. All we have to show is $\tilde{A} \in \mathscr{A}_{i+1}$. It is easy to see that there exists a set D with $\langle D \rangle = A$, $|D| \leq i$. Also, if $\tilde{A} \in \mathscr{A}_j$, then one can find a set E with $|E| \leq s - j$ and $\langle \tilde{A} \cup E \rangle = [k]$. Thus $\langle D \cup \{x\} \cup E \rangle = [k]$, giving $i + 1 + s - j \leq s$, i.e., $j \leq i + 1$. Since $|\tilde{A}| > l_i$, $j = i + 1$ follows. □

Definition 6.13. Define $b(k, L) = \max b(\mathscr{A})$, where the maximum is taken over all intersection-closed families $\mathscr{A} \subset 2^{[k]}$ with $|A| \in L$ for all $A \in \mathscr{A}$.

Conjecture 6.14 (Füredi 1983). $a(k, L) > b(k, L) - 1$ for all k and L.

Since $a(k, L) \leq b(k, L)$ by Proposition 4.17, this conjecture would mean that $\lceil a(k, L) \rceil = b(k, L)$ holds.

The smallest open cases are $L = \{0, 1, 3\}$, $k \equiv 1$ or $3 \pmod 6$, $k \geq 13$ [$b(k, L) = 3$ in this case, but $a(k, L) > 2$ is unknown for $k \neq 3^d$ or $2^d - 1$], and $L = \{0, 1, 2, 3, 5\}$, $k = 11$ [$b(k, L) = 5$ in this case]. Recently, all exponents for $k \leq 10$ were determined by Frankl et al. (1995b).

In Deza et al. (1985), an infinite family of cases where Theorem 6.3 gives the correct exponent is exhibited, e.g., $L = \{0, 1, 2, q + 1\}$, $k = q^2 + 1$, q a prime power.

For k and L with $b(k, L) = 1$, Conjecture 6.14 is obvious, since then $a(k, L) = 1$ follows from $a(k, L) \leq b(k, L)$. If $b(k; L) = 2$, then $a(k, L) > 1$ follows using constructions due to Frankl (see Füredi 1983).

A general upper bound, extending earlier results of Ray-Chaudhuri and Wilson (1975) and Babai and Frankl (1980), is the following.

Theorem 6.15 (Frankl and Wilson 1981). *Suppose that p is a prime such that $k \not\equiv l \pmod p$ holds for all $l \in L$. Let r be the number of residue classes of L modulo p. Then*

$$m(n, k, L) \leq \binom{n}{r}.$$

7. One missing intersection

An important special case of the problem treated in the preceding section is when $L = [0, k - 1] \setminus \{l\}$ for some $l \in [0, k - 1]$.

Set $m(n, k, \bar{l}) = m(n, k, [0, k - 1] \setminus \{l\})$.

There are two natural constructions for excluding the intersection size l. One is by taking all k-subsets of X containing a fixed $(l + 1)$-element subset. This gives

$$m(n, k, \bar{l}) \geq \binom{n - l - 1}{k - l - 1}.$$

The other is by taking a partial l-design. By Rödl's Theorem 6.4 this gives a lower bound of $(1-\mathrm{o}(1))\binom{n}{l}/\binom{k}{l}$. The next result of Frankl and Füredi (1985) shows that one of these constructions always gives the correct exponent.

Theorem 7.1. $m(n,k,\bar{l}) = \mathrm{O}(n^{\max\{l,k-l-1\}})$.

Proof. Consider $\mathcal{A} = \prod(\mathcal{F}^*)$ from the preceding section. We have to show that $b(\mathcal{A}) \leq \max\{l, k-l-1\}$. Let B be a base for $\mathcal{A}$ and suppose that $|B| \geq l$. For $x \in B$ consider $A_x = \langle B \setminus \{x\} \rangle \in \mathcal{A}$. Note that $A_x \cap B = B \setminus \{x\}$. Define the family (of not necessarily distinct sets)

$$\mathcal{C} = \{A_x \setminus B \colon x \in B\} \subseteq 2^{[k] \setminus B}\,.$$

Claim 7.2. *The size of the intersection of r members of $\mathcal{C}$ is never $r-c$, $1 \leq r \leq |B| = |\mathcal{C}|$, where $c = |B| - l > 0$.*

Proof. Since for distinct elements $x_1, \ldots, x_r \in B$ one has $|A_{x_1} \cap \cdots \cap A_{x_r} \cap B| = |B| - r$, $|A_{x_1} \cap \cdots \cap A_{x_r}| \neq l$ implies the claim. □

Proof of Theorem 7.1 (continued). Now a simple result of Frankl and Katona (cf. Frankl and Füredi 1985) says that any family $\mathcal{C}$ of not necessarily distinct subsets of a b-element set and satisfying the assertion of Claim 7.2 has $|\mathcal{C}| \leq b + c - 1$. Since in our case $b = k - |B|$, $c = |B| - l$, we infer that $|B| = |\mathcal{C}| \leq k - l - 1$. Since B was an arbitrary base for $\mathcal{A}$, the result follows. □

For the case $k > 2l+1$, more is true.

Theorem 7.3 (Frankl and Füredi 1985). *$m(n,k,\bar{l}) = \binom{n-l-1}{k-l-1}$ holds for $k \geq 2l+2$ and $n > n_0(k)$. Moreover, the only optimal family is $\mathcal{F} = \{F \in \binom{[n]}{k} \colon [l+1) \subset F\}$.*

For $k \leq 2l+1$ one can improve on the lower bound given by partial l-designs.

Proposition 7.4. *Let $\mathcal{P} \subset \binom{[n]}{2k-l-1}$ be a partial l-design. Then $|F \cap F'| \neq l$ for all $F, F' \in \sigma_k(\mathcal{P})$.*

Proof. Take $P, P' \in \mathcal{P}$ with $F \subset P$, $F' \subset P'$. If $P \neq P'$, then $|F \cap F'| \leq |P \cap P'| < l$. If $P = P'$ then $|F \cap F'| \geq |F| + |F'| - |P| = l+1$. □

Using Theorem 6.4 again one obtains

$$m(n,k,\bar{l}) \geq (1-\mathrm{o}(1))\binom{2k-l-1}{k}\binom{n}{l}\bigg/\binom{2k-l-1}{l}\,.$$

This inequality is partially complemented by the following result of Frankl (1983). Recall that an $S(n,a,l)$ is a partial l-design $\mathcal{S} \subset \binom{[n]}{a}$ with $|\mathcal{S}| = \binom{n}{l}/\binom{a}{l}$.

Theorem 7.5.

$$m(n, k, \bar{l}) \leqslant \binom{2k-l-1}{k}\binom{n}{k} \Big/ \binom{2k-l-1}{l}$$

holds if $k \geqslant 2l+1$ and $k-l$ is a prime power. Moreover, if $k-l$ is a prime, then equality is achieved only for $\sigma_k(\mathcal{S})$ where $\mathcal{S}$ is an $S(n, 2k-l-1, l)$.

Conjecture 7.6. Theorem 7.5 holds even if $k-l$ is not a prime power.

Settling a long-standing open problem of Erdős (cf. Erdős 1981), the following result was proved in Frankl and Rödl (1986).

Theorem 7.7. *Let $0 < \alpha \leqslant \frac{1}{4}$ and l be an integer, $\alpha n \leqslant l \leqslant (\frac{1}{2} - \alpha)n$. Then there exists $\varepsilon = \varepsilon(\alpha) > 0$ such that every family $\mathcal{F} \subset 2^{[n]}$ with $|\mathcal{F}| > (2-e)^n$ contains two sets whose intersection has size exactly l.*

For l fixed and n sufficiently large the problem was solved exactly by Frankl and Füredi (1984a). To avoid intersections of size l one can take $\mathcal{K}(n, l+1)$ which is $(l+1)$-intersecting from Katona's Theorem 5.1 and adjoin all subsets of size less than l.

Theorem 7.8. *If $\mathcal{F} \subset 2^X$ satisfies $|F \cap F'| \neq l$ for all distinct F, $F' \in \mathcal{F}$, then*

$$|\mathcal{F}| \leqslant |\mathcal{K}(n, l+1)| + \sum_{i<l} \binom{n}{i}$$

for $n > n_0(l)$.

An important tool in the proof is the following result extending Theorem 3.8 on the shadow of t-intersecting families. Recalling the definition of M, we have:

Theorem 7.9. *Suppose that the columns of $M(j, \mathcal{F})$ are linearly independent over $\mathbb{R}$, where $\mathcal{F} \subset \binom{X}{k}$. Then $|\sigma_s(\mathcal{F})|/|\mathcal{F}| \geqslant \binom{k+j}{s}/\binom{k+j}{k}$ for all $j \leqslant s < k$.*

The following problem was raised by Larman and Rogers (1972). Determine

$$s(n) = \max\{|\mathcal{F}|: \mathcal{F} \subset 2^{[n]}, |F \,\Delta F'| \neq n/2 \text{ for all } F, F' \in \mathcal{F}\}\,.$$

It is easy to see that $s(n) = 2^n$ if n is odd and that $s(n) = 2^{n-1}$ if $n \equiv 2 \pmod 4$. Let $n = 4l$ and consider the following family:

$$\mathcal{R}(l) = \{R, R^c\colon R \in 2^{[n]}, |R \cap [n-1]| \leqslant l-1\}\,.$$

Then $|\mathcal{R}(l)| = 4 \sum_{i<l} \binom{4l-1}{i}$ and $|R \,\Delta R'| \neq 2l$ for all R, $R' \in \mathcal{R}(l)$.

Theorem 7.10 (Frankl 1986a). $s(4l) = 4\sum_{i<l}\binom{4l-1}{i}$ *if* l *is the power of an odd prime.*

Conjecture 7.11. Theorem 7.9 holds for all positive integers l.

8. s-wise t-intersecting families

Let $q(n, s, t)$ denote the maximum size of an s-wise t-intersecting family $\mathscr{F} \subset 2^X$. For a more complete treatment we refer the reader to Frankl (1987b).

Proposition 8.1. $q(n, s, t)/2^{-n}$ *is monotone increasing and therefore* $q(s, t) = \lim_{n\to\infty} q(n, s, t)/2^{-n}$ *exists.*

Proof. If $\mathscr{F} \subset 2^X$ is s-wise t-intersecting, then so is $\mathscr{F}' = \mathscr{F} \cup \{F \cup \{n+1\}: F \in \mathscr{F}\}$, showing $q(n+1, s, t) \geq 2q(n, s, t)$, as desired. The second part of the proposition is a direct consequence of the first part. □

From the proposition we see that $q(s, t) \leq \frac{1}{2}$ for all $s \geq 2$, $t \geq 1$. Since $\lim_{n\to\infty} |\mathscr{H}(n, t)|/2^n = \frac{1}{2}$, $q(2, t) = \frac{1}{2}$ for all $t \geq 1$.

In view of Lemma 4.2 (iii), from now on $\mathscr{F} \subset 2^X$ will be a stable, s-wise t-intersecting family of maximum size. (Consequently, $\mathscr{F}$ is a filter.) Define the sets:

$$A_i = [n] \backslash \{t + i + ps \colon 0 \leq p \leq (n - t - i)/s\}$$

for $0 \leq i < s$ and note that

$$A_0 \cap \cdots \cap A_{s-1} = [t-1]\,. \tag{8.2}$$

Lemma 8.3. (i) $A_0 \not\in \mathscr{F}$;

(ii) *for every* $F \in \mathscr{F}$ *there exists a* $j \geq 0$ *with* $|F \cap [t + sp]| \geq t + (s-1)p$.

Proof. Since $\mathscr{F}$ is a stable filter, $A_0 \in \mathscr{F}$ would imply by repeated applications of Proposition 4.4 that $A_i \in \mathscr{F}$, $1 \leq i \leq s-1$. However, by (8.2) this is impossible, which proves (i). To prove (ii), suppose that $F = [n] \backslash \{a_0, \ldots, a_l\}$ is in $\mathscr{F}$, $1 \leq a_0 < \cdots < a_l$. If $a_p \leq t + ps$ for $0 \leq p \leq (n-t)/p$ [in particular, $l \geq (n-t)/p$], then $A_0 \in \mathscr{F}$ follows from Proposition 4.4, contradicting (i). Thus for some p we have $a_p > t + ps$, i.e., $|F \cap [t + ps]| \geq t + p(s-1)$, as desired. □

Let us consider the polynomial $x^s - 2x + 1$, for $s \geq 3$. It has exactly one root, say $\beta(s)$, in the open interval $(\frac{1}{2}, 1)$. For example, $\beta(3) = (\sqrt{5} - 1)/2$.

Theorem 8.4 (Frankl 1976). $q(n, s, t) < 2^n \beta(s)^t$.

Proof (*sketch*). Consider the probability space of all infinite $(0, 1)$-sequences with the uniform distribution. Standard computation shows that the probability of the event {there exists $p \geq 0$ such that the number of 1's up to $t + ps$ is $\geq t + p(s-1)$} is $\beta(s)^t$. By Lemma 8.3 this is a (strict) upper bound on $|\mathcal{F}|/2^n$ [we associate with $F \in \mathcal{F}$ all the (0, 1)-sequences extending its characteristic vector]. □

Define the families:

$$\mathscr{B}_p = \mathscr{B}_p(n,s,t) = \{B \subseteq [n] \colon |B \cap [t+sp]| \geq t + (s-1)p\}\,, \quad p \leq (n-t)/s\,.$$

Then $\mathscr{B}_p$ is s-wise t-intersecting and $|\mathscr{B}_p|/2^n$ is independent of n. The following result combines Theorem 8.4 and some computation involving $|\mathscr{B}_p|/2^n$.

Corollary 8.5. *There exists a positive constant c such that* $c\beta(s)^t/t < q(t,s) < \beta(s)^t$.

Conjecture 8.6. $q(n,s,t) = \max\{|\mathscr{B}_p| \colon 0 \leq p \leq (n-t)/s\}$.

Let us mention that Conjecture 8.6 holds for $s = 2$ (Katona's Theorem) and in general for $t < s \cdot 2^s/150$ (Frankl 1979). It also holds for $s \geq t \geq 2$ with $q(n,s,t) = 2^{n-t}$. Next, we show how to use this last result to give a simple proof of an important theorem of Brace and Daykin (1971).

Theorem 8.7. *Let* $\mathcal{F} \subset 2^{[n]}$ *be s-wise intersecting with* $\bigcap \mathcal{F} = \emptyset$. Then

$$|\mathcal{F}| \leq |\mathscr{B}_1(n,s,1)| = (s+2)2^{n-s-1}$$

Proof. We may suppose that $\mathcal{F}$ is a filter and thus, since $\bigcap \mathcal{F} = \emptyset$, it contains $[n]\setminus\{i\}$ for all $1 \leq i \leq n$. This will not change by shifting. Therefore, we may assume that $\mathcal{F}$ is stable.

We apply induction on s. For $s = 2$, one has $|\mathscr{B}_1(n,2,1)| = 2^{n-1}$; thus the statement follows from Theorem 1.1. Let $s \geq 3$ and suppose that Theorem 8.7 has been proved for smaller values of s. Consider $\mathcal{F}(1)$ and $\mathcal{F}(\bar{1})$.

Claim 8.8. (i) $|\mathcal{F}(1)| \leq (s+1)2^{n-s-1}$;
(ii) $|\mathcal{F}(\bar{1})| \leq 2^{n-s-1}$.

Now the theorem follows from $|\mathcal{F}| = |\mathcal{F}(1)| + |\mathcal{F}(\bar{1})|$ once we prove the claim.

Proof of Claim 8.8. Note that $\mathcal{F}(1)$ is $(s-1)$-wise intersecting on $[2,n]$, since otherwise $F_1 \cap \cdots \cap F_{s-1} = \{1\}$ for some $F_1, \ldots, F_{s-1}$ implying $\{1\} \in \bigcap \mathcal{F}$. Also, $(n]\setminus\{i\}) \in \mathcal{F}$ implies $([2,n]\setminus\{i\}) \in \mathcal{F}(1)$ for $2 \leq i \leq n$. Thus $\bigcap \mathcal{F}(1) = \emptyset$. Hence, (i) follows from the induction assumption. To prove (ii), we only have to show that $\mathcal{F}(\bar{1})$ is s-wise s-intersecting (on $[2,n]$). Otherwise, since $\mathcal{F}(\bar{1})$ is a stable filter, we can find $F_1, \ldots, F_s \in \mathcal{F}(\bar{1}) \subset \mathcal{F}$ with $F_1 \cap \cdots \cap F_s = [2,s]$. Define $G_i = (F_i\setminus\{i\}) \cup \{1\}$ for $i = 2, \ldots, s$. Then $G_i \in \mathcal{F}$ by Proposition 4.4. However, $F_1 \cap G_2 \cap \cdots \cap G_s = \emptyset$, which is a contradiction. □

9. The covering number

Recall the definition of $\tau(\mathcal{F})$.

Theorem 9.1 (Gyárfás 1977). *A k-graph $\mathcal{F}$ has at most $k^{\tau(\mathcal{F})}$ covers T of size $\tau(\mathcal{F})$.*

Proof. Set $t=\tau(\mathcal{F})$. We prove by backward induction on $l\le t$ that every l-element set is contained in at most k^{t-l} covers of $\mathcal{F}$. The case $l=t$ is trivial and the case $l=0$ will prove the theorem.

Let $0\le l<t$ and consider an l-element set A. Since $l<t=\tau(\mathcal{F})$, there exists an $F\in\mathcal{F}$ with $A\cap F=\emptyset$. Every cover of $\mathcal{F}$ containing A must contain at least one of the $(l+1)$-element sets $A\cup\{x\}$, $x\in F$. Each of these sets is (by induction) in at most k^{t-l-1} covers of $\mathcal{F}$ of size t. This gives altogether $k\cdot k^{t-l-1}=k^{t-l}$. □

For a generalization see Tuza (1988).

Considering τ pairwise disjoint sets of size k shows that Theorem 9.1 is best possible. An important corollary of the theorem is the following.

Theorem 9.2 (Erdős and Lovász 1975). *Let $\mathcal{F}$ be an intersecting k-graph with $\tau(\mathcal{F})=k$. Then $|\mathcal{F}|\le k^k$.*

Proof. Every $F\in\mathcal{F}$ is a cover of size k.

Construction (Erdős and Lovász 1975). Let $X_1,\ldots,X_k$ be disjoint sets of size $1,\ldots,k$, respectively. Define

$$\mathcal{E}_i=\{E\colon |E|=k, X_i\subset E, X_j\cap E\ne\emptyset, i<j\le k\}.$$

Set $\mathcal{E}=\mathcal{E}_1\cup\cdots\cup\mathcal{E}_k$.

Now $\mathcal{E}$ is intersecting with $\tau(\mathcal{E})=k$ and $|\mathcal{E}|=\lfloor k!e\rfloor$. Lovász conjectured that no intersecting k-graph with covering number k has more edges, but this is disproved in Frankl et al. (1995b).

How few edges can such a k-graph have?

Let $g(k)$ denote the minimum size of a k-graph $\mathcal{F}$ with $\tau(\mathcal{F})=k$. Erdős and Lovász (1975) show that $g(k)\ge 8k/3-3$ and they conjecture that $\lim_{k\to\infty} g(k)/k=\infty$. However, using an ingenious construction, Kahn (1992) proved that $g(k)=O(k)$ holds.

Let $\mathcal{P}$ be the set of lines of a projective plane of order $k-1$. Then $\mathcal{P}$ has the following strong property.

Claim 9.3. *If S is a cover of $\mathcal{P}$ with $|S|=k$, then $S\in\mathcal{P}$.*

Proof. Suppose that S is not a line and let $L \in \mathscr{P}$ be a line with $|L \cap S| \geq 2$. Choose $x \in L \backslash S$. Then there are $k-1$ lines besides L through x, and each of them has to intersect S. Thus $|S| \geq 2 + k - 1 > k$. □

Such an intersecting family is called a *maximal intersecting family*, i.e., the addition of any new k-set destroys the property of being intersecting.

Let $f(k)$ denote the minimum size of a maximal intersecting k-graph. Meyer (1974) conjectured that $f(k) \geq k^2 - k + 1$ with equality if a projective plane of order $k-1$ exists. This was disproved in Füredi (1980) by the following construction.

Example 9.4. Let $\mathscr{A}$ be the family of lines of an affine plane of order k and let $\mathscr{A} = \mathscr{L}_1 \cup \cdots \cup \mathscr{L}_{k+1}$ be the partition of the lines into parallel classes. Consider three vertex-disjoint copies $\mathscr{A}^1$, $\mathscr{A}^2$, and $\mathscr{A}^3$ of $\mathscr{A}$ and let $L_1^i, \ldots, L_k^i$ be the lines in $\mathscr{L}_{k+1}^i$. Define:

$$\mathscr{F} = \{L_j^i \cup L \colon L \in \mathscr{L}_j^{i+1}, i = 1, 2, 3, j = 1, \ldots, k\}.$$

Then $|\mathscr{F}| = 3k^2$ and $\mathscr{F}$ is a maximal intersecting family, showing $f(2k) \leq 3k^2$ if an affine plane of order k exists.

Theorem 9.5 (Boros et al. 1989). *$f(q+1) \leq q^2/2 + \mathrm{O}(q)$ for $q \equiv -1 \pmod 6$, q a prime power.*

Theorem 9.6 (Blokhuis 1987). *$f(k) \leq k^5$ for all k.*

Thus, Theorem 9.6 gives a polynomial upper bound for all k. However, it is not even known whether $\lim_{k \to \infty} f(k)/k = \infty$.

10. τ-critical k-graphs

Let us start with the following result of Bollobás (1965).

Theorem 10.1. *Let $\{A_1, \ldots, A_m\}$ and $\{B_1, \ldots, B_m\}$ be two families of subsets of $[n]$ satisfying*

(i) $A_i \cap B_i = \emptyset$, $1 \leq i \leq m$;

(ii) $A_i \cap B_j \neq \emptyset$, $1 \leq i \neq j \leq m$.

Then

$$\sum_{1 \leq i \leq m} \binom{|A_i| + |B_i|}{|A_i|}^{-1} \leq 1.$$

Proof. Apply induction on n; the cases $n = 0$, 1 are trivial. For notational convenience we shall speak of the two families as a set-pair family $\{(A_i, B_i)\colon 1 \leq i \leq m\}$ satisfying (i) and (ii).

For each $x \in [n]$ consider the set-pair family $\mathscr{P} = (A_i, B_i \setminus \{x\})$, where i runs over i with $x \not\in A_i$. Then $\mathscr{P}$ satisfies (i) and (ii). Applying the induction hypothesis to $\mathscr{P}$ on $[n] \setminus \{x\}$ and adding up the corresponding inequalities one notes that $\binom{|A_i| + |B_i|}{|A_i|}^{-1}$ occurs $n - |A_i| - |B_i|$ times, and $\binom{|A_i| + |B_i| - 1}{|A_i|}^{-1}$ occurs $|B_i|$ times. Thus we have

$$\sum_{i \leqslant i \leqslant m} (n - |A_i| - |B_i|) \cdot \binom{|A_i| + |B_i|}{|A_i|}^{-1} + |B_i| \cdot \binom{|A_i| + |B_i| - 1}{|A_i|}^{-1} \leqslant n .$$

Dividing by n, the theorem follows. □

Tuza (1984) notes that the inequality of Yamamoto (1954) is a consequence of Theorem 10.1.

Corollary 10.2. *Let $\{A_1, \ldots, A_m\}$ be an antichain on X. Then*

$$\sum_{1 \leqslant i \leqslant m} \binom{n}{|A_i|}^{-1} \leqslant 1 .$$

Proof (Tuza 1984). Set $B_i = X - A_i$ and note that the hypotheses of Theorem 10.1 are fulfilled. □

Recall the definition of τ-critical families.

Corollary 10.3. *If $\mathscr{A}$ is τ-critical with $\tau(\mathscr{A}) = t$, then*

$$\sum_{A \in \mathscr{A}} \binom{|A| + t - 1}{t - 1}^{-1} \leqslant 1 .$$

Proof. Let $\mathscr{A} = \{A_1, \ldots, A_m\}$ and let B_i be a cover of size $t - 1$ for $\mathscr{A} \setminus \{A_i\}$. Now, apply Theorem 10.1. □

Note that Corollary 10.3 implies that $|\mathscr{A}| \leqslant \binom{k+t-1}{k}$ for every τ-critical k-graph with $\tau(\mathscr{A}) = t$. Considering $\binom{[k+t-1]}{k}$ shows that this is best possible.

This result was re-proved and extended in several ways. We refer to the survey of Füredi (1988) for a full account. Here we mention only two related results.

Theorem 10.4 (Füredi 1984). *Let $(A_1, \ldots, A_m)$ be a collection of a-sets and $(B_1, \ldots, B_m)$ a collection of b-sets such that $|A_i \cap B_i| \leqslant t$ for all i and $|A_i \cap B_j| > t$ for $1 \leqslant i < j \leqslant m$. Then $m \leqslant \binom{a+b-2t}{a-t}$.*

Theorem 10.5 (Tuza 1985). *Let $\{A_1, \ldots, A_m\}$ and $\{B_1, \ldots, B_m\}$ be collections of sets with $A_i \cap B_i - \emptyset$ for all i and $(A_i \cap B_j) \cup (A_j \cap B_i) \neq \emptyset$ for $i \neq j$. Then $\sum_{1 \leqslant i \leqslant m} p^{|A_i|} q^{|B_i|} \leqslant 1$ holds for all positive p and q with $p + q = 1$.*

Proof. Let $[n]$ be the union of all the sets $A_i \cup B_i$. Consider all subsets of $[n]$ with a weight function $w(E) = p^{|E|}q^{n-|E|}$. Define $\mathscr{A}_i = \{E \subset [n]\colon\ A_i \subseteq E,\ B_i \cap E = \emptyset\}$ and note that $\mathscr{A}_1, \ldots, \mathscr{A}_m$ are pairwise disjoint. Also, note that

$$\sum_{E \in \mathscr{A}_i} w(E) = p^{|A_i|}q^{|B_i|}\,.$$

Now we can deduce the result:

$$\sum_{1 \leq i \leq m} p^{|A_i|}q^{|B_i|} = \sum_{1 \leq i \leq m} \sum_{E \in \mathscr{A}_i} w(E) \leq \sum_{E \subseteq [n]} p^{|E|}q^{n-|E|} = 1\,. \qquad \square$$

For a more general result see Tuza (1988).

11. Matchings

Let $s \geq 2$ be fixed. How large can a family $\mathscr{F} \subset 2^X$ be if $\nu(\mathscr{F}) < s$? For $s = 2$, this means that $\mathscr{F}$ is intersecting and the answer 2^{n-1} was given in Theorem 1.1.

Let $\nu(n, s)$ denote max $|\mathscr{F}|$, where $\mathscr{F} \subset 2^X$, $\nu(\mathscr{F}) < s$. Clearly, $\nu(n+1, s) \geq 2\nu(n, s)$ holds for all n. Considering $\mathscr{K} = \{K \subseteq X\colon\ |K| > n/s\}$ shows that

$$\nu(n, s) \geq \sum_{i > n/s} \binom{n}{i}\,.$$

Kleitman (1968a) showed that this is best possible for $n \equiv -1 \pmod s$.

Theorem 11.1.

$$\nu(bs-1, s) = \sum_{i=b}^{n} \binom{n}{i}\,,$$

$$\nu(bs, s) = 2\nu(bs-1, s)\,.$$

For $n \neq 0, -1 \pmod s$, the value of $\nu(n, s)$ is unknown, except for $s = 3$, where Quinn (1987) showed that for $n = 3b + 1$ the best construction is

$$\mathscr{Q} = \mathscr{K} \cup \left\{Q \in \binom{[n]}{b}\colon\ 1 \in Q\right\} = \{Q \subset [n]\colon\ |Q| + |Q \cap [1]| \geq b + 1\}\,.$$

Conjecture 11.2. For $n = bs + r$, $1 \leq r < s$,

$$\nu(n, s) = |\{K \subseteq [n]\colon\ |K| + |K \cap [s - r - 1]| \geq b + 1\}|\,.$$

A problem with a similar flavor was solved by Kleitman (1968b) for $s = 2$ and, using the same technique, by Frankl (1977) for all s.

Theorem 11.3. *Let $n = bs + s - 1$ and suppose that $\mathscr{F} \subset 2^{[n]}$ contains no s pairwise disjoint sets along with their union. Then*

$$|\mathscr{F}| \leq |G \subset [n]\colon\ b \leq |G| < bs\}|\,.$$

Again, the maximum value is unknown for $n \not\equiv -1 \pmod{s}$. Let $\nu(n,s,k)$ denote $\max|\mathscr{F}|$, where $\mathscr{F} \subset \binom{X}{k}$ and $\nu(\mathscr{F}) < s$. To avoid trivialities, suppose that $n \geqslant sk$.

Example. $\mathscr{E}_0 = \binom{[sk-1]}{k}$, $\mathscr{E}_1 = \mathscr{E}_1(n) = \{E \subset \binom{[n]}{k} \colon E \cap [s-1] \neq \emptyset\}$.

Conjecture 11.4 (Erdős 1965). $\nu(n,s,k) = \max\{|\mathscr{E}_0|, |\mathscr{E}_1|\}$.

Erdős (1965) proved that for $n > n_0(s,k)$ the conjecture is true and $\mathscr{E}_1$ is the only extremal example. Bollobás et al. (1976) show that $n_0(s,k) \leqslant 2sk^3$ holds.

The next proposition is essentially due to Kleitman (1968a).

Proposition 11.5. $\nu(ks,s,k) = \binom{ks-1}{k}$ *and, for* $s \geqslant 3$, *the only optimal family is* $\mathscr{E}_0$.

Proof. Take $\mathscr{F} \subset \binom{[ks]}{k}$ with $\nu(\mathscr{F}) \leqslant s-1$. Consider a random partition $P = (P_1, \ldots, P_s)$ of X. That is, $P_1 \cup \cdots \cup P_s = X$, $|P_i| = k$ and all P have the same chance of being chosen. Then the probability of the event $P_i \in \mathscr{F}$ is $|\mathscr{F}|/\binom{ks}{k}$. Thus the expected number of i with $P_i \in \mathscr{F}$ is $s|\mathscr{F}|/\binom{ks}{k}$. On the other hand, $\nu(\mathscr{F}) < s$ implies that this number is always less than s. Thus $s|\mathscr{F}|/\binom{ks}{k} \leqslant s-1$. [One can come to the same conclusion by the double-counting argument of Katona (1974).]

Rearranging gives $|\mathscr{F}| \leqslant \binom{ks-1}{k}$, with equality holding if and only if out of each partition P, exactly $s-1$ sets are in $\mathscr{F}$. That is, $\binom{X}{k} \backslash \mathscr{F}$ is an intersecting family of size $\binom{ks}{k} - \binom{ks-1}{k} = \binom{ks-1}{k-1}$. Now the uniqueness of $\mathscr{F}$ for $s \geqslant 3$ follows from the uniqueness part of the Erdős–Ko–Rado Theorem (see Theorem 5.3). □

Proposition 11.6. $\nu(n,s,k) \leqslant (s-1)\binom{n-1}{k-1}$ *for all* $n \geqslant sk$.

Proof. Use induction on n. The case $n = sk$ is covered by Proposition 11.5. Let $\mathscr{F} \subset \binom{X}{k}$ be a family with $|\mathscr{F}| = \nu(n,s,k)$, $\nu(\mathscr{F}) < s$. In view of Lemma 4.2 (iv) we may assume that $\mathscr{F}$ is stable. Consider the two families $\mathscr{F}(\bar{n})$, $\mathscr{F}(n)$.

Claim 11.7. $|\mathscr{F}(\bar{n})| \leqslant (s-1)\binom{n-2}{k-1}$, $|\mathscr{F}(n)| \leqslant (s-1)\binom{n-2}{k-2}$.

Since $|\mathscr{F}| = |\mathscr{F}(\bar{n})| + |\mathscr{F}(n)|$, this implies the theorem.

Proof of Claim 11.7. The first inequality is true by induction. To prove the second we have to show $\nu(\mathscr{F}(n)) < s$. Suppose the contrary and let $G_1, \ldots, G_s$ be pairwise disjoint sets in $\mathscr{F}(n)$. Since $|G_1| + \cdots + |G_s| = (k-1)s$, we can find distinct elements $x_1, \ldots, x_s \in [n] \backslash (G_1 \cup \cdots \cup G_s)$. Since $\mathscr{F}$ is stable, $G_i \cup \{x_i\}$ is in $\mathscr{F}$. That is, $\nu(\mathscr{F}) \geqslant s$, which is a contradiction. □

Formulating Proposition 11.5 for the complements $\mathscr{G} = \mathscr{F}^c$, we obtain that an s-wise intersecting family $\mathscr{G} \subset \binom{[ks]}{k(s-1)}$ can have at most $\binom{ks-1}{k(s-1)-1}$ members. This was generalized by Frankl (1976).

Theorem 11.8. *If* $\mathscr{G} \subset \binom{[n]}{l}$ *is s-wise intersecting,* $n \geq sl/(s-1)$, *then* $|\mathscr{G}| \leq \binom{n-1}{l-1}$. *Moreover, unless* $s = 2$, $n = 2l$, *equality is achieved only if* $\mathscr{G} \cong \{G \in \binom{[n]}{l}: 1 \in G\}$.

For a new proof see Frankl (1987b).

12. The number of vertices in τ- and ν-critical k-graphs

Following Tuza (1985). let us call $P = \{(A_i, B_i): 1 \leq i \leq m\}$ an (a, b)-system if $|A_i| = a$, $|B_i| = b$ for all i, and moreover, Theorem 10.1 (i) and (ii) hold.

Let $n(a, b)$ be $\max|\bigcup_{i=1}^{m} (A_i \cup B_i)|$, where the maximum is over all (a, b)-systems. Let $n_1(a, b)$ be $\max|\bigcup_{i=1}^{m} A_i|$, where the maximum is over all (a, b)-systems.

As we saw in the proof of Corollary 10.3, to every τ-critical k-graph $\mathscr{A}$ with $\tau(\mathscr{A}) = t$ one can associate a $(k, t-1)$-system. This implies:

$$|\bigcup \mathscr{A}| \leq n_1(k, t-1) \quad \text{if } \mathscr{A} \text{ is a } \tau\text{-critical } k\text{-graph with } \tau(\mathscr{A}) = t\,.$$

Obviously $n_1(a, b) \leq n(a, b) = n(b, a)$. The following surprising symmetry relation holds.

Theorem 12.1 (Tuza 1985). $n_1(a, b-1) = n_1(b, a-1)$ *for all* a, $b \geq 1$.

Proposition 12.2 (Tuza 1985).

$$n_1(a' + a'', b' + b'') \geq a' + b' + \binom{a' + b'}{a'} n_1(a'', b'')\,.$$

Proof. Let $\mathscr{P}$ (respectively, $\mathscr{Q}$) be an (a', b')-system ((a'', b'')-system). For each $(A_i, B_i) \in \mathscr{P}$, let $\mathscr{Q}_i$ be a copy of $\mathscr{Q}$, where $\mathscr{P}$, $\mathscr{Q}_1, \ldots, \mathscr{Q}_m$ are all vertex-disjoint. The general element of $\mathscr{Q}_i$ is denoted by $(C_j^{(i)}, D_j^{(i)})$. Define:

$$\mathscr{R} = \{(A_i \cup C_j^{(i)}, B_i \cup D_j^{(i)}): (A_i, B_i) \in \mathscr{P}, (C_j^{(i)}, D_j^{(i)}) \in \mathscr{Q}_i\}\,.$$

Then $\mathscr{R}$ is an $(a' + a'', b' + b'')$-system, which proves the theorem. □

Tuza (1985) proves the following surprisingly sharp bounds.

Theorem 12.3.

(i) $$\frac{1}{4}\binom{a+b+1}{b+1} < n(a, b) < \binom{a+b+1}{b+1} \quad \textit{for } a \geq b, a \geq 1;$$

(ii) $$\frac{1}{4}\binom{a+b+1}{b+1} < n_1(a, b) < \binom{a+b+1}{b+1} \quad \textit{for } a \geq 1, b \geq 0\,.$$

Let us mention that Tuza proves both the upper and lower bounds in a stronger form. In particular, applying Proposition 12.2 with $a' = \lfloor ab/(b+1) \rfloor$, $b' = b$, he obtains

$$n_1(a,b) \geq \lceil a/(b+1) \rceil \binom{\lfloor ab/(b+1) \rfloor + b}{b} + \lfloor ab/(b+1) \rfloor + b$$

and he suggests that equality holds here for $a \geq b+2$. He also conjectures that $n_1(a,b) = n(a,b)$ holds if and only if $a \geq b$.

Recall the definition of a ν-critical family $\mathscr{F}$. A family $\mathscr{F}$ is said to have *rank* k if $k = \max_{F \in \mathscr{F}} |F|$ holds.

Improving earlier bounds of Lovász (1975), Tuza (1985) shows:

Theorem 12.4. *If $\mathscr{F}$ is a ν-critical family of rank k, then it has fewer than $\binom{\nu(\mathscr{F})k+k}{k}$ vertices.*

Proof. Set $\nu = \nu(\mathscr{F})$ and let $\mathscr{H}$ consist of those sets which are the union of ν pairwise disjoint edges in $\mathscr{F}$. Let $\mathscr{H}' = \{H_1, \ldots, H_m\} \subset K$ be minimal with respect to $\bigcup \mathscr{H}' = \bigcup \mathscr{H}$. Then for every $H_i \in \mathscr{H}'$ there is a vertex $x_i \in H_i$ such that $x_i \not\in H_j$ for $i \neq j$. By ν-criticality there is some $F_i \in \mathscr{F}$ with $F_i \cap H_i = \{x_i\}$ and consequently $(F_1 \backslash \{x_i\}) \cap H_j \neq \emptyset$ for all $i \neq j$. Now $\{(H_i, F_i \backslash \{x_i\})\colon 1 \leq i \leq m\}$ is a system satisfying Theorem 10.1 (i) and (ii), and also $|H_i| \leq \nu k$, $|F_i \backslash \{x_i\}| \leq k-1$. Thus,

$$\left|\bigcup \mathscr{F}\right| = \left|\bigcup \mathscr{H}'\right| \leq n_1(\nu k, k-1) < \binom{\nu k + k}{k}. \qquad \square$$

In the case $\nu = 1$ we have the following sharper results.

Theorem 12.5 (Tuza 1985). *Let $\nu(k)$ denote the maximum order of a ν-critical intersecting family $\mathscr{F}$ with rank k. Then*

$$2k - 4 + 2\binom{2k-4}{k-2} \leq \nu(k) \leq \binom{2k-1}{k-1} + \binom{2k-3}{k-2}.$$

Both bounds improve earlier results of Erdős and Lovász (1975). Tuza conjectures that the lower bound – given by the following construction – is optimal for $k \geq 4$.

Example. For each partition $[2k-4] = F \cup F'$ with $|F| = |F'| = k-2$, take four new vertices x, x', y, y' and form the k-element sets $F \cup \{x, y\}$, $F \cup \{x', y'\}$, $F' \cup \{x, y'\}$ and $F' \cup \{x', y\}$. These sets form a ν-critical k-graph.

For k fixed and ν large we have the following.

Conjecture 12.6 (Lovász 1975). There exists a constant $c = c(k)$ such that every ν-critical family $\mathscr{F}$ of rank k has at most $c\nu(\mathscr{F})$ vertices. For $k = 2$, the best possible bound $3\nu(\mathscr{F})$ was shown by Gallai (1963).

13. Excluded configurations I

Let $\mathscr{C} = \{\mathscr{A}_1, \ldots, \mathscr{A}_r\}$ be a collection of k-graphs.

Set $\mathrm{ex}(n, \mathscr{C}) = \max|\mathscr{F}|$, where the maximum is taken over all $\mathscr{F} \subseteq \binom{X}{k}$, $\mathscr{F}$ containing no subfamily isomorphic to a family in $\mathscr{C}$. If $\mathscr{C} = \{\mathscr{A}\}$ then we also write $\mathrm{ex}(n, \mathscr{A})$ instead of $\mathrm{ex}(n, \mathscr{C})$. A classical result of Katona et al. (1964) is the following.

Theorem 13.1. *$\mathrm{ex}(n, \mathscr{C})/\binom{n}{k}$ is monotone decreasing, and therefore $\mu(\mathscr{C}) = \lim_{n\to\infty} \mathrm{ex}(n, \mathscr{C})/\binom{n}{k}$ exists.*

Proof. Let $1 \leqslant h < n$ and consider a family $\mathscr{F} \subset \binom{X}{k}$ without any subfamily isomorphic to some $\mathscr{A} \in \mathscr{C}$ and such that $|\mathscr{F}| = \mathrm{ex}(n, \mathscr{C})$. Choose a subset $H \in \binom{X}{h}$ at random, with uniform distribution. Then $|\binom{H}{k} \cap \mathscr{F}| \leqslant \mathrm{ex}(h, \mathscr{C})$ holds for all H. On the other hand, the expectation of $|\binom{H}{k} \cap \mathscr{F}|$ is $|\mathscr{F}|$ times the probability that a fixed $F \in \binom{X}{k}$ is in $\binom{H}{k}$, i.e., $|\mathscr{F}|\binom{h}{k}/\binom{n}{k}$. Thus $|\mathscr{F}|\binom{h}{k}/\binom{n}{k} = \mathrm{ex}(n, \mathscr{C})\binom{h}{k}/\binom{n}{k} \leqslant \mathrm{ex}(h, \mathscr{C})$.

Dividing by $\binom{h}{k}$ shows the desired result. □

It follows from a result of Erdős (1964) that $\mu(\mathscr{C}) = 0$ if $\mathscr{C}$ contains some k-partite k-graph. Actually, Erdős obtains an upper bound of the form $n^{k-e(\mathscr{C})}$ where $e(\mathscr{C})$ is a positive constant. The determination of the best possible value of $e(\mathscr{C})$ seems to be very difficult even in very simple cases. In this section we suppose that there are no k-partite k-graphs in $\mathscr{C}$. Let us first state Turán's well-known problem.

Example. Let $[n] = X_0 \cup X_1 \cup X_2$ be a partition with $|X_i| = \lfloor (n+i)/3 \rfloor$. Define:

$$\mathscr{T}(4,3) = \left\{T \in \binom{X}{3} \colon |T \cap X_i| = 1, i = 0, 1, 2\right\}$$
$$\cup \left\{T \in \binom{X}{3} \colon |T \cap X_i| = 2, |T \cap X_{i+1}| = 1 \text{ for some } i = 0, 1, 2, \right.$$
$$\left. \text{where } X_3 \text{ denotes } X_0 \right\}.$$

It is conjectured by Turán that $t(n, 4, 3) = |\mathscr{T}(4,3)|$. Kostoschka (1982) has given exponentially many non-isomorphic 3-graphs with $|\mathscr{T}(4,3)|$ edges and without a $\binom{[4]}{3}$. This suggests, that if Turán's conjecture is correct, then it could be very hard to prove. Kalai (1985) has proposed a more general algebraic conjecture.

Example. Let $[n] = X_0 \cup X_1 \cup \cdots \cup X_{t-1}$ be a partition with $|X_i| = \lfloor (n+i)/t \rfloor$.

Define:

$$\mathcal{T}(n, t(k-1)+1, k) = \binom{[n]}{k} - \bigcup_{0 \le i < t} \binom{X_i}{k}.$$

Clearly, $\mathcal{T}(n, t(k-1)+1, k)$ contains no $\binom{[t(k-1)+1]}{k}$. It is conjectured that for $n > n_0(t, k)$ one has $|\mathcal{T}(n, t(k-1)+1, k)| = t(n, t(k-1)+1, k)$, although Brown (1983) has produced other examples with the same cardinality.

The simplest non-3-partite 3-graph is $\mathcal{R}_3 = \binom{[4]}{3} \setminus [2, 4]$. Even for this 3-graph, $\mathrm{ex}(n, \mathcal{R}_3)$ is unknown.

Proposition 13.2. $\frac{2}{7} \le \mu(\mathcal{R}_3) \le \frac{1}{3}$.

Here, the upper bound is due to de Caen (1982), the lower bound to Frankl and Füredi (1984b).

With every k-graph $\mathcal{F}$ let us associate a polynomial $q(\mathcal{F})$ as follows.

Definition 13.3. Define $q(\mathcal{F}, x) = \sum_{F \in \mathcal{F}} \prod_{i \in F} x_i$.

Then $q(\mathcal{F})$ is a homogeneous polynomial of degree k which is linear in each variable.

Define the *Lagrange function* $\lambda(\mathcal{F}) = \max q(\mathcal{F}, x)$, where the maximum is then over all $x = (x_1, \ldots, x_n)$ with $x_i \ge 0, x_1 + \cdots + x_n = 1$.

Using the theory of Lagrange multipliers one obtains:

Lemma 13.4 (Frankl and Rödl 1984). *There exists an* $x = (x_1, \ldots, x_n)$ *with* $x_i \ge 0$, $x_1 + \cdots + x_n = 1$, *such that* (i)–(iii) (*following*) *hold. Set* $Y = \operatorname{supp} x = \{i\colon x_i > 0\}$.

(i) $\lambda(\mathcal{F}) = q(\mathcal{F}, x)$;

(ii) $\partial q(\mathcal{F}, x)/\partial x_i = k\lambda(\mathcal{F})$ *for all* $i \in Y$;

(iii) *every pair* $P \in \binom{Y}{2}$ *is contained in some edge* $F \in \mathcal{F}$ *with* $F \subseteq Y$.

Note that $\lambda(\mathcal{F}) \ge |\mathcal{F}|/n^k$. One can use this to show the following simple result:

$$\mu(\mathcal{C}) = k! \sup\{\lambda(\mathcal{F})\colon \mathcal{F} \text{ is a } k\text{-graph without a copy of any } \mathcal{A} \in \mathcal{C}\}. \tag{13.5}$$

Katona (1974) asked for the determination of the maximum number $\mathrm{symm}(n, k)$ of k-subsets of an n-set such that none of them contains the symmetric difference of two others. This problem can be formulated in terms of $\mathrm{ex}(n, \mathcal{C})$, but for k large $\mathcal{C}$ will contain many k-graphs (all with three edges).

Conjecture 13.5 (Bollobás 1974). $\mathrm{symm}(n, k) = \prod_{0 \le i < k} \lfloor (n+i)/k \rfloor$ with equality holding for the complete equipartite k-graph.

Bollobás (1974) solves the case $k = 3$ (the case $k = 2$ is very easy and was solved by Mantel in 1906). De Caen (1986) gives a new proof for $k=3$ and proposes a different problem.

Problem. Determine $\mathrm{ex}(n, \mathscr{C}_k) = c(n,k)$, where $\mathscr{C}_k = \{\mathscr{A}_2, \mathscr{A}_3, \ldots, \mathscr{A}_k\}$ with $\mathscr{A}_i = \{[1,k], [1,k-1] \cup \{k+1\}, [i, k+i-1]\}$.

Clearly, $\mathrm{symm}(n,k) \leqslant c(n,k)$ for all k and for $k=2$, 3, the two problems are the same.

Sidorenko (1987) realized the relevance of Lemma 13.4 and proved the following.

Theorem 13.6. *$c(n,k) = \prod_{0 \leqslant i < k} \lfloor (n+i)/k \rfloor$ holds for $k = 2, 3, 4$.*

Proof. To avoid technical difficulties we shall only prove $c(n,k) \leqslant (n/k)^k$ (which is the same as the theorem if n is a multiple of k), i.e., $\lambda(\mathscr{F}) \leqslant 1/k^k$ if $\mathscr{F}$ contains no copy of $\mathscr{A}_i \in \mathscr{C}_k$.

In view of Lemma 13.4 in proving the above inequality we may suppose that $\sigma_2(\mathscr{F}) = \binom{[n]}{2}$, i.e., every air $P \in \binom{[n]}{2}$ is contained in some $F \in \mathscr{F}$. Now if F, $F' \in \mathscr{F}$ with $|F \cap F'| = k-1$, then $|F \Delta F'| = 2$ and therefore we find $F'' \in \mathscr{F}$ with $F \Delta F' \subseteq F''$, i.e., $\{F, F', F''\} \in \mathscr{F}$, which is a contradiction. Thus $|F \cap F'| \leqslant k-2$ for all F, $F' \in \mathscr{F}$. (Note that this is not true in general for $\mathscr{F}$ containing no copy of $\mathscr{A}_i \in \mathscr{C}_k$; however, Lemma 13.4 ensures the existence of a subfamily with this property and the same value for the Lagrange function.) That is, $\binom{F}{k-1} \cap \binom{F'}{k-1} = \emptyset$ for distinct F, $F' \in \mathscr{F}$. In other words, $\partial q(\mathscr{F}, x)/\partial x_i$ and $\partial q(\mathscr{F}, x)/\partial x_j$ have no common term for $i \neq j$. Let

$$s_{k-1}(x) = \sum_{A \in \binom{[n]}{k-1}} \prod_{i \in A} x_i$$

be the $(k-1)$th elementary symmetric polynomial. Adding up Lemma 13.4 (ii) for $1 \leqslant i \leqslant n$, we obtain

$$kn\lambda(\mathscr{F}) \leqslant s_{k-1}(x) \leqslant \binom{n}{k-1}\left(\frac{1}{n}\right)^{k-1}.$$

Rearranging gives

$$\lambda(\mathscr{F}) \leqslant \frac{(n-1)\cdots(n-k+2)}{k!n^{k-1}}. \tag{13.7}$$

Now for $n \geqslant k$, the right-hand side of (13.7) is at most k^{-k}, both for $k=2$ and $k=3$, and also for $k=4$ unless $n=5$. However, the case $n=5$ is impossible, because any two 4-subsets of [5] overlap in three elements. This concludes the proof. □

Using the same approach, Frankl and Füredi (1989) determined $\mu(\mathscr{C}_k)$ for $k=5$ and $k=6$.

Let $\mathcal{W}_{11}$ ($\mathcal{W}_{12}$) be the (unique) (11, 5, 4) ((12, 6, 5)) Steiner-system. That is, $\mathcal{W}_{12} \subset \binom{[12]}{6}$ and for each $A \in \binom{[12]}{5}$ *there is a unique set* $B \in \mathcal{W}_{12}$ with $S \subset B$, and $\mathcal{W}_{11} = \mathcal{W}_{12}\,(12)$.

Example. For $X = X_1 \cup \cdots \cup X_{12}$, $|X_i| = n/12$, define:

$$\mathcal{B}_6 = \left\{ B \in \binom{X}{6} \colon \{i\colon\ B \cap X_i \neq \emptyset\} \in \mathcal{W}_{12} \right\};$$

$\mathcal{B}_5$ is defined analogously.

Theorem 13.8. (i) $\mathrm{ex}(n, \mathscr{C}_5) \leq 66(n/11)^5$ *with equality iff* $11 \mid n$, *in which case* $\mathcal{B}_5$ *is the only optimal family.*

(ii) $\mathrm{ex}(n, \mathscr{C}_6) \leq 132(n/12)^6$ *with equality iff* $12 \mid n$, in which case $\mathcal{B}_6$ *is the only optimal family.*

14. Excluded configurations II: k-partite k-graphs

Many of the problems treated earlier can be formulated in the form: determine $\mathrm{ex}(n, \mathscr{C})$. For example, the determination of $m(n, k, L)$ is such a problem. Let us start with three problems which come up in other contexts.

Call a family $\mathcal{F} \subset 2^X$ *barely overlapping if* $F \not\subseteq F' \cup F''$ holds for all distinct F, F', $F'' \in \mathcal{F}$. Let $h(n, k)$ denote the maximum size of a barely overlapping family $\mathcal{F} \subset \binom{X}{k}$.

Theorem 14.1 (Erdős et al. 1982). (i) $h(n, 2l-1) \leq \binom{n}{l}/\binom{2l-1}{l}$ *with equality holding iff there exists an* $S(n, 2l-1, l)$.

(ii) $h(n, 2l) \leq \binom{n-1}{l}/\binom{2l-1}{l}$ *with equality achieved for some* $\mathcal{F}$ *if and only if* $|\bigcap \mathcal{F}| = 1$ (say $\bigcap \mathcal{F} = \{1\}$) and $\mathcal{F}(1)$ is an $S(n-1,\ 2l-1, l)$.

Proof. We only prove (i) and even this only for $n \geq 3l$. Let $G \subset F \in \mathcal{F}$. We call G a *distinguished subset* of F if $G \not\subseteq F'$ for all $F \neq F' \in \mathcal{F}$. Let us define a weight function $w : \mathcal{F} \times \binom{[n]}{l} \to \mathbb{R}_+$ by:

$$w(F, G) = \begin{cases} 1 & \text{if } G \in \binom{F}{l} \text{ and } G \text{ is an eigen-subset of } F\,, \\ 1/l & \text{if } G \cap F =: H \in \binom{F}{l-1} \text{ and } H \text{ is an eigen-subset of } F\,, \\ 0 & \text{otherwise}\,. \end{cases}$$

Claim. $\sum_F w(F, G) \leq 1$, $\sum_G w(F, G) \geq \binom{2l-1}{l}$.

Proof. The first part follows by noting that if G is an eigen-subset of F, then no subset of G can be an eigen-subset of some other $F' \in \mathcal{F}$ and $w(F, G) = 1/l$ can hold for a fixed G at most l times, once for each of its $(l-1)$-subsets

To prove the second part, note that if $F = A \cup B$, $|A| = l$, $|B| = l-1$, then either A or B (or both) are eigen-subsets of F because $\mathcal{F}$ is barely overlapping. If A is an eigen-subset, it contributes 1; if B is, then B contributes $(n-(k-1))/l >$ 1. Since there are $\binom{2l-1}{l}$ such partitions of F, the inequality follows. □

Proof of Theorem 14.1 (continued). Using the claim, it is easy to show that

$$|\mathcal{F}|\binom{2l-1}{l} \leq \sum_{F\in\mathcal{F}} \sum_{G\in\binom{X}{l}} w(F,G) = \sum_{G}\sum_{F} w(F,G) \leq \binom{n}{l},$$

i.e., $|\mathcal{F}| \leq \binom{n}{l}/\binom{2l-1}{l}$, as desired. In case of equality, equality must hold in (i). Thus all $G \in \binom{F}{l}$ are eigen-subsets. That is, $\mathcal{F}$ is a partial l-design. Consequently, $|\mathcal{F}| = \binom{n}{l}/\binom{2l-1}{l}$ if and only if $\mathcal{F}$ is an $S(n, 2l-1, l)$. □

For further results and problems on barely overlapping and related families we refer to Frankl (1988).

We call $\mathcal{F} \subset 2^X$ *union-free* if $F \cup F' = G \cup G'$ implies for $F, F', G, G' \in \mathcal{F}$ that $\{F, F'\} = \{G, G'\}$. Let $u(n,k)$ denote $\max|\mathcal{F}|$, where $\mathcal{F} \subset \binom{X}{k}$ is union-free.

Theorem 14.2 (Frankl and Füredi 1986a). *There are positive constants c_k, c'_k such that*

$$c_k n^{2k/3+\varepsilon(k)} < u(n,k) < c'_k n^{2k/3+\varepsilon(k)},$$

with $\varepsilon(k) = 0$, $\frac{1}{3}$ or $\frac{1}{6}$ according to whether $k \equiv 0$, 1 or 2 (mod 3).

Let us mention that the proof of the lower bound is rather involved. The acquired family $\mathcal{F}$ is defined via systems of nonlinear equations over finite fields.

Again, for more information on this and related problems we refer to Frankl (1988).

We call $\mathcal{F}$ disjoint-union-free if it contains no four sets F, G, H, K with $F \cup G = H \cup K$ and $F \cap G = H \cap K = \emptyset$. Let $u_d(n,k)$ denote the maximum size of $\mathcal{F} \subset \binom{X}{k}$, $\mathcal{F}$ disjoint-union-free.

Clearly, $u_d(n,k) \geq \binom{n-1}{k-1} + 1$; (take $\{G \in \binom{[n]}{k}\colon 1 \in G\} \cup \{[2, k+1]\}$). It is possible that for $k \geq 4$, $n > n_0(k)$, equality holds. However, it was unknown for many years whether $u_d(n,k) = O(n^{k-1})$ held. Füredi (1983) gave an ingenious argument to show the following.

Theorem 14.3. $u_d(n,k) < \frac{7}{2}\binom{n}{k-1}$ *for all $n > k \geq 3$.*

Let us note that in the case of graphs ($k = 2$), the condition is that the graph is C_4-free and $u_d(n, 2)$ is of the order $n^{3/2}$ (cf. chapter 23).

The paper by Frankl and Füredi (1987) gives a rather general treatment (and often solutions) of a class of excluded-configuration-type problems for k-graphs. We mention just a few of the results of that paper.

Let $\phi(a, b)$ be the maximum size of an a-graph without sunflowers of size b. Also, let $\phi(n, k, l, s)$ be the maximum size of $\mathscr{F} \subset \binom{X}{k}$, where $\mathscr{F}$ contains no sunflower of size s whose center has size l.

Theorem 14.4. $\phi(n, k, l, s) = (\phi(l+1, s) + o(1))\binom{n-l-1}{k-l-1}$ *if* $k > 2l + 1$.

It is conjectured that the same holds for $k = 2l + 1$ as well; however, this has only been proved (in Chung and Frankl 1987) for $l = 1$. For $k < 2l + 1$ it follows from Theorem 7.1 via Lemma 4.15 that $\phi(n, k, l, s)$ has order n^l; however, the correct coefficient of n^l is unknown.

Conjecture 14.5.

$$\phi(n, k, l, s) = \left(\binom{l-1+s(k-l)}{k} + o(1)\right)\binom{n}{l} \Big/ \binom{l-1+s(k-l)}{l} .$$

The construction is given by taking $\mathscr{F} = \sigma_k(\mathscr{S})$, where $\mathscr{S} \subset \binom{X}{l+1+s(k-l)}$ is a (partial) l-design.

Let $\mathscr{A}$ be a k-graph. Set $p = |\bigcap \mathscr{A}|$ and let q be the number of vertices of $\mathscr{A}$ of degree 2 or more.

Theorem 14.6. *If* $2p + q + 1 < k$, *then* $\mathrm{ex}(n, \mathscr{A}) = (\gamma(\mathscr{A}) - o(1))\binom{n-p-1}{k-p-1}$, *where* $\gamma(\mathscr{A})$ *is a positive integer depending only on* $\mathscr{A}$.

In the case $p = 0$, one can define $\gamma(\mathscr{A})$ by taking $\gamma(\mathscr{A}) + 1$ to be the size of the smallest set T satisfying $|T \cap A| = 1$ for all $A \in \mathscr{A}$. Note that such a T exists if $\mathscr{A}$ is k-partite, which – in turn – follows from $q < k$. In general, $\gamma(\mathscr{A}) \le \phi(p+1, |\mathscr{A}|)$.

Theorem 14.7. *Set* $\mathscr{A} = \{\{1,2,3,5,7\}, \{1,2,3,6,8\}, \{1,2,4,5,8\}\}$. *Then* $\mathrm{ex}(n, \mathscr{A}) = o(n^4)$. *However,* $\lim_{n\to\infty} \mathrm{ex}(n, \mathscr{A})/n^\alpha = \infty$ *for all* $\alpha < 4$.

This result shows that $\mathrm{ex}(n, \mathscr{A})$ does not always have a proper exponent. The proof extends that of Ruzsa and Szemerédi (1978), where a similar phenomenon is described.

Another type of extremal problem, considered by Kászonyi and Tuza (1986), is the following.

Definition 14.8. For a k-graph $\mathscr{A}$, let $\mathrm{sat}(n, \mathscr{A})$ denote $\min |\mathscr{F}|$, where $\mathscr{F} \subset \binom{[n]}{k}$, and $\mathscr{F}$ contains no copy of $\mathscr{A}$, but adding any new k-subset of $[n]$ produces a copy of $\mathscr{A}$.

Conjecture 14.9 (Tuza). $\mathrm{sat}(n, \mathscr{A}) = O(n^{k-1})$ for every k-graph $\mathscr{A}$.

References

Bang, C., H. Sharp and P. Winkler

[1984] On coverings of a finite set: depth and subcover, *Period. Math. Hungar.* **15**, 51–60.

Blokhuis, A.

[1987] More on maximal intersecting families of finite sets, *J. Combin. Theory A* **44**, 299–303.

Blumer, A., A. Ehrenfeucht, D. Hausler and M.K. Warmuth

[1989] Learnability and the Vapnik–Chervonenkis dimension, *J. ACM* **36**, 929–965.

Bollobás, B.

[1965] On generalized graphs, *Acta Math. Acad. Sci. Hungar.* **16**, 447–452.

[1974] Three graphs without two triples whose symmetric difference is contained in a third, *Discrete Math.* **8**, 21–24.

Bollobás, B., D.E. Daykin and P. Erdős

[1976] Sets of independent edges of a hypergraph, *Quart. J. Math. Oxford* **27**, 25–32.

Boros, E., Z. Füredi and J. Kahn

[1989] Maximal intersecting families and affine regular polygons in PG$(2,q)$, *J. Combin. Theory A* **52**, 1–9.

Brace, A., and D.E. Daykin

[1971] Cover theorems for finite sets, *Bull. Aust. Math. Soc.* **5**, 197–202.

[1972] Sperner type theorems for finite sets, in: *Proc. British Combinatorial Conf., Oxford* (Institute of Mathematics and its Applications, Southend-on-Sea) pp. 18–37.

Brown, W.G.

[1983] On an open problem of P. Turán concerning 3-graphs, in: *Studies in Pure Mathematics to the Memory of P. Turán*, ed. P. Erdős, pp. 91–93.

Cameron, P.J., P. Frankl and W.M. Kantor

[1989] Intersecting families and permutation groups without fixed-point-free 2-elements, *European J. Combin.* **10**.

Chung, F.R.K., and P. Frankl

[1987] The maximum number of edges in a 3-graphs not containing a given star, *Graphs and Combin.* **3**, 111–126.

Clarkson, K., H. Edelsbrunner, L. Guibas, M. Sharir and E. Welzl

[1988] Combinatorial complexity bounds for arrangements of curves and surfaces, in: *Proc. 29th Annu. Symp. on Comput. Geom.*, pp. 568–599.

de Caen, D.

[1982] *On Turán's hypergraph problem*, Ph.D. Thesis (University of Toronto).

Deza, M., P. Erdős and P. Frankl

[1978] Intersection properties of systems of finite sets, *Proc. London Math. Soc.* (3) **36**, 368–384.

Deza, M., P. Frankl and J.W.P. Hirschfeld

[1985] Sections of varieties over finite fields as large intersection families, *Proc. London Math. Soc.* **50**, 405–425.

Engel, K., and H.D.O.F. Gronau

[1985] *Sperner Theory of partially ordered sets, Teubner Texte in Mathematik* (Teubner, Leipzig).

Erdős, P.

[1964] On extremal problems on graphs and generalized graphs, *Israel J. Math.* **2**, 183–190.

[1965] A problem of independent r-tuples, *Ann. Univ. Budapest* **8**, 93–95.

[1981] On the combinatorial problems which I would most like to see solved, *Combinatorica* **1**, 25–42.

Erdős, P., and D.J. Kleitman

[1968] On colorings of graphs to maximize the proportion of multicolored k-edges, *J. Combin. Theory* **5**, 164–169.

Erdős, P., and L. Lovász

[1975] Problems and results on 3-chromatic hypergraphs and some related questions, *Colloq. Math. Soc. János Bolyai*, pp. 609–627.

Erdős, P., and R. Rado

[1960] Intersection theorems for systems of sets, *J. London Math. Soc.* **35**, 95–90.

Erdős, P., C. Ko and R. Rado

[1961] Intersection theorems for systems of finite sets, *Quart. J. Math. Oxford (2)* **12**, 313–320.

Erdős, P., P. Frankl and Z. Füredi

[1982] Families of finite sets in which no set is covered by the union of two others, *J. Combin. Theory A* **33**, 158–166.

Frankl, P.

[1975] The proof of a conjecture of Katona, *J. Combin. Theory A* **19**, 208–213.

[1977] Families of finite sets containing no k sets and their union, *Period. Math. Hungar* **8**, 29–31.

[1978] The Erdős-Ko-Rado theorem is true for $n = ckt$, *Colloq. Math. Soc. János Bolyai* **18**, 365–375.

[1983] On the trace of finite sets, *J. Combin. Theory A* **34**, 41–45.

[1984] A new short proof for the Kruskal-Katona theorem, *Discrete Math.* **48**, 327–329.

[1986a] Orthogonal vectors and codes with missing distances, *Combinatorica* **6**, 279–286.

[1986b] All rationals occur as exponents, *J. Combin. Theory A* **42**, 200–206.

[1987a] Erdős-Ko-Rado theorem with conditions on the maximal degree, *J. Combin. Theory A* **46**, 252–263.

[1987b] The shifting technique in extremal set theory, in: *Combinatorial Surveys 1987*, ed. C. Whitehead (Cambridge University Press, Cambridge) pp. 81–110.

[1988] Intersection and containment problems without quantitative restrictions, in: *Algebraic, Extremal and Metric Combinatorics* (Cambridge University Press) pp. 62–111.

Frankl, P., and Z. Füredi

[1981] A short proof for a theorem of Harper about Hamming spheres, *Discrete Math.* **34**, 311–313.

[1984a] On hypergraphs without two edges intersecting in a given number of vertices, *J. Combin. Theory A* **36**, 230–236.

[1984b] An exact result for 3-graphs, *Discrete Math.* **50**, 323–328.

[1985] Forbidding just one intersection, *J. Combin. Theory A* **39**, 160–176.

[1986] Union free families of sets and equations over fields, *J. Number Theory* **23**, 210–218.

[1987] Exact solution of some Turán-type problems, *J. Combin. Theory A* **45**, 226–262.

[1989] Extremal problems whose solutions are the blow-ups of the small Witt-designs, *J. Combin. Theory A* **52**, 129–147.

Frankl, P., and V. Rödl

[1984] Hypergraphs do not jump, *Combinatorica* **4**, 149–159.

[1985] Near perfect coverings in graphs and hypergraphs, *European J. Combin.* **6**, 317–326.

[1987] Forbidden intersections, *Trans. Amer. Math. Soc.* **300**, 259–286.

Frankl, P., and R.M. Wilson

[1981] Intersection theorems with geometric consequences, *Combinatorica* **1**, 357–368.

Frankl, P., K. Ota and N. Tokushige

[1995a] Exponents of uniform L-systems, *J. Combin. Theory A*, submitted.

[1995b] Uniform intersecting families with covering restrictions, *J. Combin. Theory A*, to appear.

Frankl, P., M. Matsumoto, I.Z. Ruzsa and N. Tokushige

[1995c] Minimum shadows in uniform hypergraphs and a generalization of the Takagi function, *J. Combin. Theory A* **69**, 125–148.

Füredi, Z.

[1980] On maximal intersecting families of finite sets, *J. Combin. Theory A* **28**, 282–289.

[1983] On finite systems whose every intersection is a kernel for a star, *Discrete Math.* **47**, 129–132.

[1984] Geometrical solution of an intersection problem for two hypergraphs, *European J. Combin.* **5**, 133–136.

[1988] Extremal problems concerning the principal parameters of hypergraphs, *Graphs and Combin.* **4**, 115–206.

Füredi, Z., and J.R. Griggs

[1986] Families of finite sets with minimum shadows, *Combinatorica* **6**, 393–402.

Gallai, T.

[1963] Neuer Beweis eines Tutteschen Satzes, *MTA Mat. Kut. Int. Közl*, 135–139.

Gyárfás, A.

[1977] Partition covers and blocking sets in hypergraphs (in Hungarian), *MTA SZTAKI Tanulmányok* 71, Budapest.

Harper, L.H.

[1966] Optimal numberings and isoperimetric problems on graphs, *J. Combin. Theory* **1**, 385-394.

Hilton, A.J.W.

[1976] unpublished Manuscript.

Hilton, A.J.W., and E.C. Milner

[1967] Some intersection theorems for systems of finite sets, *Quart. J. Math. Oxford (2)* **18**, 369-384.

Kahn, J.

[1992] On a problem of Erdős and Lovász, *Combinatorica* **12**, 417-423.

Kalai, G.

[1985] A new approach to Turán's conjecture, *Graphs and Combin.* **1**, 107-109.

Kászonyi, L., and Z. Tuza

[1986] Saturated graphs with minimal number of edges, *J Graph Theory* **10**, 203-210.

Katona, G.O.H.

[1964] Intersection theorems for systems of finite sets, *Acta Math. Hungar.* **15**, 329-337.

[1966] A theorem of finite sets, in: *Theory of Graphs, Proc. Colloq., Tihany* (Akadémiai Kiadó, Budapest) pp. 187-207.

[1974] A simple proof of the Erdős-Ko-Rado theorem, *J. Combin. Theory B* **13**, 183-184.

Katona, G.O.H., T. Nemetz and M. Simonovits

[1964] On a problem of Turán in the theory of graphs, *Mat. Lapok* **15**, 228-238 (in Hungarian, English summary).

Kleitman, D.J.

[1966a] Families of non-disjoint subsets, *J. Combin. Theory* **1**, 153-155.

[1966b] On a combinatorial conjecture of Erdős, *J. Combin. Theory* **1**, 209-214.

[1968a] Maximal number of subsets of a finite set no k of which are pairwise disjoint, *J. Combin. Theory* **5**, 157-163.

[1968b] On families of subsets of a finite set containing no two disjoint sets and their union, *J. Combin. Theory* **5**, 235-237.

Kostoshka, A.V.

[1982] A class of constructions for Turán (3, 4) problem, *Combinatorica* **2**, 187-192.

Kruskal, J.B.

[1963] The number of simplices in a complex, in: *Mathematical Optimization Techniques* (University of California Press, Berkeley, CA) pp. 251-278.

Larman, D.G., and C.A. Rogers

[1972] The realization of distances within sets in Euclidean space, *Mathematika* **19**, 1-24.

Linial, N., Y. Mansour and R.L. Rivest

[1991] Results on learnability and the Vapnik-Chervonenkis dimension, *Information and Computation* **90**, 33-49.

Lovász, L.

[1975] On minimax theorems of combinatorics (in Hungarian), *Matematikai Lapok* **26**, 209-264.

[1979] *Combinatorial Problems and Exercises* (Akadémiai Kiadó/North-Holland, Budapest/Amsterdam).

Matsumoto, M., and N. Tokushige

[1989] The exact bound in the Erdős-Ko-Rado Theorem for cross-intersecting families, *J. Combin. Theory A* **22**, 90-97.

Meyer, J.C.

[1974] Quelques problémes concernant les cliques des hypergraphes h-complets et q-parti h-complets, in: *Hypergraph Seminar, Columbus, Ohio, 1972, Lecture Notes in Mathematics*, Vol. 411, eds. C. Berge et al. (Springer, Berlin) pp. 127-139.

Miklós, D.

[1984] Great intersecting families of edges in hereditary hypergraphs, *Discrete Math.* **48**, 95-99.

Mörs, M.
[1985] A generalization of a theorem of Kruskal, *Graphs and Combin.* 1, 167–183.

Quinn, F.
[1987] Ph.D. Thesis (MIT, Cambridge, MA).

Rödl, V.
[1985] On a packing and covering problem, *European J. Combin.* 6, 69–78.

Ruzsa, I.Z., and E. Szemerédi
[1978] Triple systems with no six points carrying three triangles, *Colloq. Math. Soc. János Bolyai* **18**, 939–945.

Sidorenko, F.
[1987] On a problem of Bollobás on 4-graphs (in Russian), *Mat. Zametki* **41**, 433–455.

Sperner, E.
[1928] Ein Satz über Untermengen einer endlichen Menge, *Math. Z.* **27**, 544–548.

Tuza, Z.
[1984] Helly-type hypergraphs and Sperner families, *European J. Combin.* 5, 185–187.
[1985] Critical hypergraphs and intersecting set-pair systems, *J. Combin. Theory B* **39**, 134–145.
[1988] *An Inequality for Minimal Covering Sets in Set Systems of Given Rank*, Preprint (MTA SzTAKI, Budapest).

Wilson, R.M.
[1984] The exact bound in the Erdős–Ko–Rado theorem, *Combinatorica* 4, 247–257.

Yamamoto, K.
[1954] Logarithmic order of free distributive lattices, *J. Math. Soc. Jpn.* 6, 343–353.

There are many aspects of combinatorics which are mirrored by chapters in this Handbook. But few areas form such a compact body of concepts and results (and thus in turn form a theory in the classical sense) as Ramsey theory. When presenting it one could follow the axiom–definition–theorem–corollary scheme.

We decided otherwise and we follow roughly an historical account commented from a contemporary point of view. It seems to us that this is the most appropriate method for the subject.

Our aim is to give a survey of the recent development of Ramsey theory and include proofs of some of the main results. Recently found techniques enable us to do so.

Very few areas of combinatorics display such a variety of techniques from various parts of mathematics. This should not be seen as a surprise. Many results of Ramsey theory (including Ramsey's theorem itself) have a character of a combinatorial principle which may be viewed as a structural generalization of the pigeon-hole principle. Often such principles mirror profound features of objects studied in various areas of mathematics.

The first version of this paper was written in 1988 and since then it has been updated several times. It is advantageous that we can build upon several books and surveys which are devoted either to various aspects of Ramsey theory (such as Graham 1981, Graham et al. 1980, Nešetřil and Rödl 1978a, 1979b, 1990a, Rödl 1991, Prömel and Voigt 1990) or contain chapters devoted to our subject (Bollobás 1985, 1978, 1979, Alon and Spencer 1992). However, no previous knowledge of Ramsey theory is necessary.

I wish to thank Hillel Fürstenberg and Vojtech Rödl for valuable discussions and generous help.

1. Ramsey's theorem

We start with the result which gave the subject its name.

Finite Ramsey theorem 1.1 (Ramsey 1930). *Let p, t, n be positive integers. Then there exists a positive integer N with the following property*:

> *If X is a set with at least N elements and $a_1 \cup \cdots \cup a_t$ is any partition of the set $\binom{X}{p}$ of all p-element subsets of X, then there exists a subset Y of X with at least n elements such that the set $\binom{Y}{p}$ is a subset of one of the classes a_i of the partition.* (1.1)

Admittedly this is a technical statement and several helpful notions and convenient notations will be introduced. Moreover, it is a characteristic of Ramsey's theorem (and Ramsey theory) that it often occurs hidden or in a different context as might be suggested by the following notions and notations:

The set $\{1, \ldots, n\}$ will be denoted by $[n]$ and the set $\{Y\colon Y \subseteq X \text{ and } |Y| = p\}$ by $\binom{X}{p}$. This convenient notation originated in the context of Ramsey theory, and it is due to Leeb (1973). The partition $a_1 \cup \cdots \cup a_t$ (in the above statement of Ramsey's theorem) is called a *coloring*; it is sometimes written as $c: \binom{X}{p} \rightarrow [t]$. We shall see that the most important case is $t = 2$. The set Y is called a *homogeneous set*, or sometimes a *monochromatic set*.

The validity of statement (1.1) (for particular values n, p, t and N) is traditionally denoted by the symbol

$$N \rightarrow (n)^p_t$$

which is called the Erdős–Rado partition arrow (see Erdős 1987 or chapter 42). The smallest value of N for which $N \rightarrow (n)^p_t$ holds will be denoted by $r(p, t, n)$ and is called *Ramsey number*.

As a warm up let us list some particular cases: It is easy to see that $r(1, t, n) = t(n - 1) + 1$. This is the Anglo-American pigeonhole and continental Schubfach- (also Dirichlet's) principle. The case $p = 2$ and $t = 2$ is the most frequent case – the *graph case*: every graph with at least N vertices satisfies either $\alpha(G) \geqslant n$ or $w(G) \geqslant n$ (i.e. every sufficiently large party contains either n mutual acquaintances or n mutual strangers). In this setting $r(2, t, 3) - 1$ is the maximum number of vertices in a complete graph which can be decomposed into t triangle-free graphs. Very very few Ramsey numbers are known exactly. We shall return to this aspect, which is perhaps the typical feature for the whole field, in section 3.

Contemporary proofs of Ramsey's theorem do not differ much from the classical proofs. We give two proofs which are similar; nevertheless the mild formalism of the second extends to much more general situations.

Proof I. We proceed by induction on p (for every choice of t, n). The case $p = 1$ we discussed above. In the induction step let the existence of numbers $r(p - 1, t, n)$ be proved for every choice of t and n. Fix $t > 1$ and $n > p$ to avoid trivial cases. We prove the existence of $r(p, t, n)$ by giving an upper bound for it.

First put $t_0 = r(1, t, n)$ and for $k = 1, \ldots, t_0 - 1$ define numbers r_k by

$$r_{k+1} = r(p - 1, t, r_k) + 1$$

with $r_1 = 1$.

We claim that $r(p, t, n) \leqslant r_{t_0} = r$. To prove this let X be a set of size r and let $c: \binom{X}{p} \rightarrow [t]$ be an arbitrary coloring. Without loss of generality assume that $X = \{1, \ldots, r\}$. Proceeding by backward induction for $k = t_0, t_0 - 1, \ldots, 1$ define sets $X = X_{t_0} \supseteq \cdots \supseteq X_1$, $|X_k| = r_k$, and elements $1 = x_{t_0}, \ldots, x_1$ as follows: If X_k is defined, let x_k be the minimal element of X_k. Put $X'_k = X_k - \{x_k\}$ and define the *induced* coloring c'_k of the set $\binom{X'_k}{p}$ by $c'_k(A) = c(A \cup \{x_k\})$. By the construction of the number r_k there exists a set $X_{k-1} \subseteq X'_k$ of size r_{k-1} which is homogeneous (with respect to c'_k); denote such a set by X_{k-1}. This definition guarantees that $X_1 \neq \emptyset$. Now consider the set $X' = \{x_{t_0}, x_{t_0 - 1}, \ldots, x_1\}$ and define the (induced) coloring $c': X' \rightarrow \{1, \ldots, t\}$ by

$$c'(x_k) = c(\{x_k\} \cup A)$$

where A is any $(p-1)$-element subset of $\{x_{k-1}, \ldots, x_1\}$. (The above backward induction just guarantees that the coloring c' is well-defined.) Finally, using the choice of t_0, let $Y \subseteq X'$ be a c'-homogeneous set with n points. We claim that Y is homogeneous. However, this is clear: for if P, $P' \in \binom{Y}{p}$ then

$$c(P) = c'(\min P) = c'(\min P') = c(P) . \qquad \square$$

Proof II. Let us recast the above proof of Ramsey's theorem in different language: By a set we mean now a set of positive integers. The *beginning* of a set is its minimal element. If X is a set then a coloring of $\binom{X}{p}$ is said to be *good* if any two sets with the same beginning have the same color.

Ramsey's theorem follows immediately from the following two claims.

Claim A (Sufficiency of good colorings). *The following two statements are equivalent (for a fixed p and t):*

(1) *For any positive integer n there exists a positive integer N with the following property: If X is a set with at least N elements together with an arbitrary t-coloring of $\binom{X}{p}$ then there exists an n-subset Y of X such that the coloring restricted to $\binom{Y}{p}$ is good; we denote this by*

$$N \underset{\text{good}}{\longrightarrow} (n)^p_t ;$$

(2) *For any positive integer n there exists N such that $N \rightarrow (n)^p_t$.*

Claim B (Existence of good colorings). *For every n there exists N such that $N \underset{\text{good}}{\longrightarrow} (n)^p_t$.*

Proof of Claim A is easy since any set Y of size $t(n-1)+1$ for which the coloring of $\binom{Y}{p}$ is good contains an n-subset which is monochromatic.

Claim B can be proved easily by induction on n using Ramsey's theorem for $p-1$: Namely if $N \underset{\text{good}}{\longrightarrow} (n)^p_t$ then

$$\bar{N} = 1 + r(p-1, t, N) \underset{\text{good}}{\longrightarrow} (n+1)^p_t .$$

One simply considers all p-sets which contain 1 and finds a set $X \subseteq \{2, \ldots, \bar{N}\}$ of size N with all p-sets of form $\{1\} \cup Z$, $Z \in \binom{X}{p-1}$, monochromatic. $\square$

Admittedly Proofs I and II are similar with a few cosmetic changes. However, Ramsey's theorem is a part of combinatorics where the choice of notions ("language") matters. Below we shall benefit from this formulation of Ramsey's theorem by giving a strikingly similar proof of the Graham–Rothschild theorem (see section 4).

F.P. Ramsey discovered this theorem in a sound mathematical context. Perhaps because of this, Ramsey's theorem was never regarded as a puzzle and combinatorical curiosity. It's beauty and power are now well established. However, it was largely through the efforts of P. Erdős that the subject enjoys the current high level of popularity and research activity. Erdős together with G. Szekeres initiated the applications of Ramsey's theorem in geometry by proving the following (Erdős and Szekeres 1935).

Theorem 1.2. *Let n be a positive integer. Then there exists an N with the following property: If X is a set of at least N points in the plane no three of which are collinear then X contains an n-tuple Y which forms the vertices of an convex n-gon.*

(Classical hint: Prove $N \leqslant r(4,5,n)$.)

Apart from the problem of a good estimation of the optimal value of N (which is a common "difficulty" with all Ramsey-type results) there is a peculiar structural problem here:

Call a set $Y \subseteq X$ an *n-hole* in X if Y is the set of vertices of a convex n-gon which does not contain other points of X.

Problem 1.3. Does there exist $\mathcal{N}(n)$ such that if X is any set of at least $\mathcal{N}(n)$ points in the plane no three of which are collinear then X contains an n-hole?

It is easy to prove that $\mathcal{N}(n)$ exists for $n \leqslant 5$ (see Harborth 1978); Horton (1983) showed that $\mathcal{N}(n)$ does not exist for $n \geqslant 7$. The existence of $\mathcal{N}(6)$ is open. See Valtr (1992) and Nešetřil and Valtr (1994) for a recent discussion of this problem.

An easy variant of the above proofs of Ramsey's theorem lead to the bound

$$r(p,k,n) \leqslant 2^{2^{\cdot^{\cdot^{\cdot^{2^{c_k m}}}}}}$$

where c_k is a positive constant and the stack of 2's has height $p-1$. In Ramsey theory we frequently meet very large numbers (and the theory is the source of very large cardinals, called *Erdős cardinals*; see (Erdős et al. 1965, 1984a or chapter 42). To estimate some of these numbers is sometimes cumbersome which is in the present style resolved by saying "assume that N is sufficiently large". The meaning of the phenomena of these "large" functions took recently an unexpected turn in the context of mathematical logic and the notions of primitive recursive functions and provably (total) functions invaded finite Ramsey theory. We shall touch this in section 3.3.

Let us return for a moment to the origins of Ramsey theory.

It has often been said that although Ramsey discovered the theorem in a meaningful mathematical context, the later development obscured this motivation in the favor of the combinatorial part (of his unique (!)) mathematical paper. As a result of this wide belief, the logical part of his paper is even omitted from a collection of his works (Ramsey 1978). But one never knows. Recently, exactly this part of Ramsey's research found a very nice application. Let us briefly mention it.

Ramsey's paper contains a sophisticated application of "Ramsey's theorem" to the *decision problem* for a class of first-order formulas. Explicitly, Ramsey proves that if φ is a first-order formula of the form

$$\exists x_1 \exists x_2 \cdots \exists x_n \forall y_1 \cdots \forall y_m \Phi ,$$

where Φ is quantifier-free (such formulas are called *Bernays–Schönfinkel formulas*) then there exists an algorithm which decides whether φ holds for every finite structure. This seemed to be a dead end project as (shortly after Ramsey's death) it was shown (by Church) that the decision problem is algorithmically unsolvable, indeed, for first-order formulas (Trakhtenbrot 1950) and even for first-order formulas of the form

$$\exists x_1 \cdots \exists x_n \forall y_1 \cdots \forall y_m \exists z_1 \cdots \exists z_p \Phi ,$$

see Lewis (1979).

However, the whole area got a new turn when Glebski et al. (1969) and, independently, Fagin (1976) showed that for every first-order formula φ, the probability $P(A, \varphi)$ of the fact that a formula φ is satisfied by a structure A with n points either tends to 1 or tends to 0 as n tends to infinity (see chapter 6).

This result (and similar ones) is called a 0–1 *law* (here for first-order formulas). However, Glebski's et al. result does not cover all instances. In fact the most interesting properties are related to the properties described by the second-order formulas, e.g., Hamiltonicity is defined by a second-order formula: "there exists an *ordering* of the vertices which forms a cycle"; P (Hamiltonicity) tends to 1 by a result of Posa (1976), see chapter 6 of this volume.

Efforts were made to extend the 0–1 law to classes of special second-order sentences. (For second-order formulas in general the problem is unsolvable.) Kolaitis and Vardi (1987) proved that the class of second-order formulas derived from Bernays–Schönfinkel formulas with only one existential second-order quantifier with its first-order part being Bernays–Schönfinkel, satisfies the 0–1 law and they determined the complexity of the corresponding decision problem. It is possible to say that their proof uses ideas of Ramsey's original paper with only one "slight" change: one has to replace Ramsey's theorem with a more modern device, namely the Structural Ramsey theorem, proved by Nešetřil and Rödl in (1977a), see section 5.

It is amazing how persistent the original motivations are.

Since this introductory section is devoted mostly to classical papers, let us mention that the earlier of the widely known theorems of Ramsey-type is the following result due to Schur.

Theorem 1.4 (Schur 1916). *For every t there exists a positive integer N such that for every partition of the set $\{1, \ldots, N\}$ into t classes, one of the classes contains numbers x and y together with their sum $x + y$.*

Proof. Schur's theorem easily follows from Ramsey's theorem. Take $N = r(2, t, 3)$. Let $c: \{1, \ldots, N\} \rightarrow \{1, \ldots, t\}$ be a fixed coloring. Define then the coloring c' of pairs by $c'(\{i, j\}) = c(|i - j|)$. By the choice of N there exists a

c'-monochromatic triangle with vertices $i<j<k$. However, then $c(i-j)=c(k-j)=c(k-i)$ and $(j-i)+(k-j)=k-i$. □

It is fair to say that Schur has been a hidden force behind the development of Ramsey theory in the thirties: he was the first one to formulate the conjecture which Van der Waerden turned into theorem (see Van der Waerden 1927 and the next section). His student R. Rado soon thereafter in 1933 obtained one of the deepest results of Ramsey theory (see Rado 1933).

The motivation for Schur's research in this area was algebraic (p-adic versions of Fermat's problem and the distribution of quadratic residues).

2. Adventures of arithmetic progressions

Besides Ramsey's theorem itself the following result provided constant motivation for the theory.

Theorem 2.1 (Van der Waerden 1927). *For every choice of positive integer t and n there exits N such that for every partition of the set $\{1,2,\ldots,N\}$ into t classes one of the classes contains an arithmetic progression with n terms.*

The original proof of Van der Waerden (which arose in a discussion with Artin and Schreier, see Van der Waerden (1971) for the account of the discovery) and which is included in the enchanting and moving book of Chintschin (1951), was until recently principally the only known proof. However, interesting modifications of the proof were also found (see, e.g., Graham and Rothschild 1974, Deuber 1982, and Taylor 1982). Most important of these is probably the combinatorial formulation of the Van der Waerden result found by Hales and Jewett (1963).

Theorem 2.2 (Cube theorem; Hales–Jewett theorem). *Let A be a finite set (alphabet) and let t be a positive integer. Then there exists N with the following property: For every partition of the cube A^N into t classes one of the classes contains a combinatorial line.*

Here we think of A^N as the set of all vectors $(x_1,\ldots,x_N)$ with each of its entries belonging to A (i.e., an N-dimensional *cube* over A).

A *combinatorial line* is a set of points of A^N of the following form:

$$\{(x_1,\ldots,x_N): x_i=x_j \text{ for } i,j\in I, x_i=x_i^0 \text{ for } i\not\in I\},$$

where $(x_1^0,\ldots,x_N^0)$ is a fixed point of A^N and I is a non-empty subset of $\{1,\ldots,N\}$.

The Hales–Jewett theorem readily implies Van der Waerden's theorem: If we put $A=\{0,1,\ldots,n-1\}$ and with a point $x=(x_1,\ldots,x_N)$, we associate an integer

$$w(x) = \sum_{i=1}^{N} x_i \cdot n^i$$

then the mapping $w: A^N \to Y$ is 1–1 and it is easy to see that every combinatorial line is mapped to an arithmetic progression of length n.

On the other hand, as mentioned earlier the original proof of the Hales–Jewett theorem is closely related to Van der Waerden's proof and in fact it may be viewed as its combinatorial axiomatization.

However, the distinctive feature of both proofs is that one has to prove a more general statement and then to use a double induction. This procedure does not provide a primitive recursive upper bound for the size of N: Denote by $W(n)$ the minimal number N which satisfies to Van der Waerden's theorem (for $t = 2$). Known values are $W(2) = 3$, $W(3) = 9$, $W(4) = 35$, $W(5) = 178$. The upper bound for $W(n)$ supplied by the original proofs grows like $A(n)$, where A is the Ackermann function. The Ackermann function may be defined by the following procedure.

Ackermann hierarchy 2.3. For each positive integer n, define the function

$$f_n : \mathbb{N} \to \mathbb{N}$$

as follows:

$$f_1(i) = i + 1\,,$$
$$f_2(i) = 2 \cdot i\,,$$
$$f_{n+1}(i) = \underbrace{f_n \circ \cdots \circ f_n}_{i}(1)\,.$$

(Thus $f_3(i) = 2^i$ and $f_4(i)$ is a stack of 2's of height i (the tower function).) Thus, Ramsey numbers $r(p, k, n)$ are bounded by the function f_4.

The Ackermann function A is the diagonal function

$$A(n) = f_n(n)\,.$$

(A fails to be primitive recursive; it cannot be expressed by a combination of the usual function operations).

On the other hand, the best lower bound (for n prime) is (only!) $W(n + 1) > n2^n$ (due to Berlekamp 1968). Thus, the question of whether such a huge upper bound was necessary, was one of the main research problems in this area.

This feeling was not new and already Erdős and Turán (1936), for the purposes of improving the estimates for $W(n)$, got the idea of trying to prove a stronger (now called a *density*) statement: Denote by $r_k(n)$ the smallest r such that an arbitrary sequence

$$1 \leq a_1 < \cdots < a_r \leq n$$

must contain an arithmetic progression of length k. Clearly a good upper bound on the function $r_k(n)$ implies Van der Waerden's theorem and it may yield a better bound on $W(n)$.

However, this approach proved to be very difficult and so far did not contribute to the bounding of $W(n)$. Yet it proved to be a very rich source of problems and results. Even the basic question whether

$$r_k(n) = \mathrm{o}(n)$$

appeared to be very difficult. Let us remark that because of the (clear) subadditivity

$$r_k(m+n) \leq r_k(m) + r_k(n)$$

the limit

$$\lim r_k(n)/n \rightarrow r_k$$

exists. It was a stunning achievement that the problem was solved affirmatively by Szemerédi (1975) fifty years later. We state his result in the finite form as follows.

Theorem 2.4 (Szemerédi's theorem). *For every $\varepsilon > 0$ and positive integer n, there exists N such that every subset X of $\{1, 2, \ldots, N\}$ of size at least εN contains an arithmetic progression with n terms.*

Szemerédi's proof is purely combinatorial and very complicated. Several pieces proved to be useful in other contexts, most notably, his so-called Regularity lemma (Szemerédi 1976), see chapter 23.

Szemerédi's regularity lemma has many beautiful applications to Ramsey theory, see, e.g., Nešetřil and Rödl (1979b), Chvátal et al. (1983) or Chen and Schelp (1993) and recent Ajtai et al. (1994) or Erdős et al. (1995). We shall mention these on several places below. However a great achievement, it is true that Szemerédi's proof is a mathematical *tour de force*. Thus great excitement was caused 15 years ago by a new proof of Szemerédi's theorem by means of *ergodic theory*. This has been found by Fürstenberg (1977), see his monography (Fürstenberg 1981). It is beyond the scope of this chapter to describe these methods. Let us just add a few remarks.

Ergodic theorems deal with infinite structures (and yield the finite Ramsey type theorems). In order to start this work one has to properly generalize Ramsey type statements to infinite sets. This is straightforward for the Ramsey theorem itself (see Theorem 3.4) but for number theoretical results less so. It seems that the topological dynamics provides a proper non-elementary setting for many Ramsey type (coloring) questions in combinatorial number theory by systematically using the global properties of structures (such as automorphisms of $\mathbb{Z}$), see Fürstenberg and Weiss (1978), Hindman (1979), Carlson and Simpson (1990) where this is made explicit (see also section 4.4). Ergodic methods profoundly extend and use the topological dynamics results to yield density results. This comment is strengthened by a very recent development (the density version of the Hales–Jewett theorem, see section 4.3). The ergodic methods enabled Fürstenberg, Katznelson and Weiss to prove various results which were previously too complex or even inaccessible for combinatorial methods. For example one can prove both Van der Waerden's theorem and Szemerédi's theorem for higher-dimensional lattices). Whereas Van der Waerden's theorem for n-dimensional lattices is not difficult to derive from the 1-dimensional case (this is known as the Gallai–Witt theorem, see section 4.4) a similar generalization of the Szemerédi's theorem has been proved first by ergodic theory means and no combinatorial proof is presently known.

Another example are the so-called *iterated density theorems*: Observe that neither for Schur's theorem nor for Ramsey's theorem an analogy of Szemerédi's theorem holds (for integers from $n/2$ to n form a sum-free set, and a complete bipartite graph with n vertices may have a positive proportion of the edges of the complete graph with n vertices and yet contains no triangle). However, given a set X of integers one may consider occurrences of these x for which there are "many" y such that $x+y$ is also in the set. This led to the notion of iterated density theorem which was considered and initiated from the point of view of topological dynamics by Bergelson (1986), and Bergelson and Hindman (1988), and from the combinatorial point of view in a series of papers by Frankl et al. (1988). (However, combinatorial methods, sometimes seem to yield stronger results here.)

One last comment on topological dynamics methods: The basic notion in transforming combinatorial number theory to topological dynamics theory is that of a dynamical system which is a compact metric space Y together with a homeomorphism $T: Y \rightarrow Y$. In combinatorial applications Y is induced by (say) 2-colorings of the set $\mathbb{Z}$ of all integers and T is the shift operator defined by the shift of the coloring to the right: $(Tx)(i) = x(i+1)$.

All topological dynamics versions of combinatorial number-theoretical results involve this interpretation using shift operators and perhaps this led Bergelson to formulate the following conjecture: Let $H = (\mathbb{Z}, \mathcal{M})$ be a hypergraph such that the shift $i \rightarrow i+1$ is an automorphism of $(\mathbb{Z}, \mathcal{M})$. Suppose the chromatic number is infinite. Denote by α_n the maximal independent set contained in the subhypergraph induced by the set $\{1, \ldots, n\}$. Then α_n/n tends to 0 (as n tends to infinity).

If true this conjecture would imply Szemerédi's theorem. However, it was disproved by Ruzsa and, independently, by Kříž (1987) who disproved it even for graphs.

All of the above mentioned methods are existential in nature and they do not produce any bounds for the Van der Waerden numbers $W(n)$. The problem of the upper bound for the function $W(n)$ was generally felt as one of the main problems of the field and the situation raised speculation relating the problem to the hierarchies of rapidly growing functions (upon which we shall comment in section 3). Thus, the work of Shelah in this area had the effect of a bombshell. Shelah (1988) found a new proof of Van der Waerden's theorem and even the Hales–Jewett theorem which avoids use of double induction and yields a primitive recursive upper bound. The method of the proof is remarkably simple and we give it here:

In the proof we shall use the following technical lemma.

Shelah's pigeonhole lemma 2.5. *For all positive integers n and t, there exists m with the following property: For every choice of colorings $c_l : m^{2n-1} \rightarrow [t]$, $l = 1, 2, \ldots, n$, there are integers $1 \leq a_l < b_l \leq m$ such that for every $l = 1, \ldots, m$, we have*

$$\begin{aligned} &c_l(a_1, b_1, \ldots, a_{l-1}, b_{l-1}, a_l, a_{l+1}, b_{l+1}, \ldots, a_n, b_n) \\ &\quad = c_l(a_1, b_1, \ldots, a_{l-1}, b_{l-1}, b_l, a_{l+1}, b_{l+1}, \ldots, a_n, b_n) . \end{aligned} \tag{2.1}$$

Denote by $f(n, t)$ the minimal such m.

Proof. We proceed by induction on n. Obviously $f(1, t) = t + 1$. For the inductive step we prove

$$f(n+1, t) \leq t^{f(n,t)^{2n}} + 1 = N . \tag{2.2}$$

Put $M = f(n, t)$. Let $c_l : [N]^{2n+1} \rightarrow [t], l = 1, \ldots, n+1$, be arbitrary colorings. Consider the induced coloring

$$c' : [N] \rightarrow [t]^{[M]^{2n}}$$

defined by

$$c'(a) = (c_{n+1}(x_1, \ldots, x_{2n}, a); (x_1, \ldots, x_{2n}) \in [M]^{2n}) .$$

By the pigeonhole principle there are $1 \leq a_{n+1} < b_{n+1} \leq N$ with

$$c'(a_{n+1}) = c'(b_{n+1}) . \tag{2.3}$$

Now we can define t-colorings $c''_l : [M]^{2n-1} \rightarrow [t]$, $l = 1, \ldots, n$, by

$$c''_l(x_1, \ldots, x_{2n-1}) = c_l(x_1, \ldots, x_{2n-1}, a_{n+1}, b_{n+1}) .$$

By the induction hypothesis there are numbers

$$a_1 < b_1, a_2 < b_2, \ldots, a_n < b_n$$

such that for every $l = 1, \ldots, n$, we have

$$\begin{aligned} &c''_l(a_1, b_1, \ldots, a_{l-1}, b_{l-1}, a_l, a_{l+1}, b_{l+1}, \ldots, a_n, b_n) \\ &\quad = c''_l(a_1, b_1, \ldots, a_{l-1}, b_{l-1}, b_l, a_{l+1}, b_{l+1}, \ldots, a_n, b_n) \\ &\quad = c_l(a_1, b_1, \ldots, a_{l-1}, b_{l-1}, b_l, a_{l+1}, b_{l+1}, \ldots, a_n, b_n, a_{n+1}, b_{n+1}) \\ &\quad = c_l(a_1, b_1, \ldots, a_{l-1}, b_{l-1}, a_l, a_{l+1}, b_{l+1}, \ldots, a_n, b_n, a_{n+1}, b_{n+1}) . \end{aligned}$$

But (2.3) implies

$$c_{n+1}(a_1, b_1, \ldots, a_n, b_n, a_{n+1}) = c_{n+1}(a_1, b_1, \ldots, a_n, b_n, b_{n+1})$$

which all together give (2.1). □

Proof of the Hales–Jewett theorem. Denote by $\mathrm{HJ}(n, t)$ the minimal number N for which the statement of the Hales–Jewett theorem (for $A = [n]$ and t colors) holds.

Obviously $\mathrm{HJ}(1, t) = 1$ and by induction on n we prove

$$\mathrm{HJ}(n + 1, t) \leq \mathrm{HJ}(n, t) \cdot f(\mathrm{HJ}(n, t), t^{(n+1)^{\mathrm{HJ}(n,t)}}) , \tag{2.4}$$

where f is the function occurring in Shelah's pigeonhole lemma 2.5.

To simplify the notation, set $N = \mathrm{HJ}(n, t)$ and $m = f(N, t^{(n+1)^N})$. (Thus the desired upper bound (2.4) is $N \cdot m$.) Also set $M_l = \{ml + 1, \ldots, m(l + 1)\}$, for $l = 1, \ldots, N$.

Now let $c: [n + 1]^{N \cdot m} \to [t]$ be a fixed coloring.

In fact, it suffices to consider a coloring of a subset of $[n + 1]^{N \cdot m}$ formed by all cascade functions: A *cascade function* f determined by a family $(a_l, b_l$: $l = 1, \ldots, N)$, $a_l \leq b_l$, a_l, $b_l \in M_l$, and a function $g \in [n + 1]^N$ is a function belonging to $[n + 1]^{N \cdot m}$ which satisfies

$$f(i) = \begin{cases} n + 1 & \text{for } i < a_l, i \in M_l , \\ g(l) & \text{for } a_l \leq i \leq b_l , \\ n & \text{for } b_l < i \in M_l . \end{cases}$$

$(a_l, b_l$: $l = 1, \ldots, N)$ is called the *schema* S of the cascade function f.

For a fixed schema S the mapping $g \mapsto f$ is a 1–1 mapping which carries a line in $[n + 1]^N$ into a line in $[n + 1]^{N \cdot m}$. We put $f = H_S(g)$. (See the schematic fig. 2.1.)

Let us define colorings

$$d_l: [m]^{2N-1} \to \{f ; f: [n + 1]^N \to [t]\}$$

by

$$d_l(a_1, b_1, \ldots, a_{l-1}, b_{l-1}, a_l = b_l, a_{l+1}, b_{l+1}, \ldots, a_N, b_N) \\ = (c(H_{(a_l, b_l)}(g)); g \in [n + 1]^N) \tag{2.5}$$

(if (a_l, b_l) does not correspond to a schema of a cascade function then we define d_l arbitrarily).

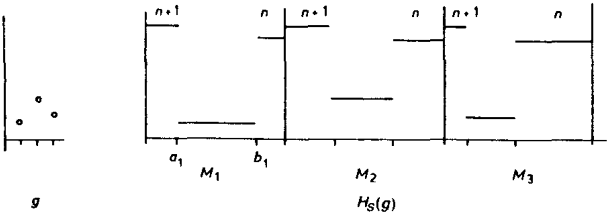


Figure 2.1.

By Shelah's pigeonhole lemma there exists a schema S

$$S = \langle a_1 < b_1, a_2 < b_2, \ldots, a_N < b_N \rangle$$

such that (2.1) holds; explicitly, for every $l = 1, \ldots, N$,

$$\begin{aligned} &d_l(a_1, b_1, \ldots, a_{l-1}, b_{l-1}, a_l, a_{l+1}, b_{l+1}, \ldots, a_N, b_N) \\ &\quad = d_l(a_1, b_1, \ldots, a_{l-1}, b_{l-1}, b_l, a_{l+1}, b_{l+1}, \ldots, a_N, b_N) . \qquad (2.6) \end{aligned}$$

Let us consider all cascade functions with schema S, i.e., all functions $H_S(g)$, $g \in [n+1]^N$. By the choice $N = \mathrm{HJ}(n, t)$, there exists $I \subseteq [N]$ and $x^0 = (x_1^0, \ldots, x_N^0)$ such that if we denote by L the line in $[n]^N$ determined by x^0 and I then $H_S(L)$ is a monochromatic subset of $[n+1]^{N \cdot m}$.

Let the points (functions) of the line L be denoted by $x^1, \ldots, x^n$. Observe that all the cascade functions $H_S(x^1)$, $H_S(x^2), \ldots, H_S(x^n)$ have schema S. Define $x^{n+1} = (x_l^{n+1}, \ldots, x_N^{n+1})$ by

$$x_i^{n+1} = \begin{cases} x_i^0 & \text{for } i \notin I , \\ n+1 & \text{for } i \in I . \end{cases}$$

The cascade functions $H_S(x^{n+1})$ has schema S. However, both cascade functions $H_S(x^n)$ and $H_S(x^{n+1})$ may be also thought of as having any of the following schemas $((a'_l, b'_l)\colon l = 1, \ldots, N)$ where $a'_l = a_l, b'_l = b_l$, for $l \notin I$, and

$$a'_l, b'_l \in \{a_l, b_l\} , \quad a'_l \leq b'_l \quad \text{for } l \in I .$$

From this follows (by repeated use of (2.5) and (2.6)) that

$$c(H_S(x^n)) = c(H_S(x^{n+1}))$$

and thus $x^1, x^2, \ldots, x^{n+1}$ is a monochromatic line in $[n+1]^{N \cdot m}$. □

Corollary 2.6.

$$\mathrm{HJ}(n, t) \leq f_5(c \cdot (n+t)) ,$$

where f_5 is the function introduced in the Ackerman hierarchy.

Proof. The function $f(n, t)$ has a growth of the tower function which in our notation is the function f_4. The bound on $\mathrm{HJ}(n, t)$ then involves the iteration of $f(n, t)$ as the principal term which gives the function f_5. □

This corollary justifies the subtitle of Shelah's preprint: indeed "from the Ackerman sphere into the atmosphere"; well, perhaps, stratosphere.

In the spirit of P. Erdős, R.L. Graham recently offered \$1000 for the proof of an f_3-type upper bound for the Van der Waerden function $W(n)$. For a very compact write up of Shelah's proof, see Nilli (1990). Another version appears in the second edition of Graham et al. (1980).

Let us add one remark. The Hales–Jewett theorem found an effective use in many proofs of structural extensions of Ramsey's theorem (see section 6). Thus we have a primitive recursive bound for all these results. Particularly, this is true for Ramsey's theorem for set systems (or structural Ramsey theorem) and Ramsey's theorem for space systems which we state later in section 5. However, the strongest results in this area fail to fall presently into this category since their proofs use double induction.

To end this chapter let us return once again to the Van der Waerden theorem from a more number-theoretical point of view:

The most frequently studied case of the Erdős and Turán numbers $r_k(n)$ is, of course, $k = 3$.

Here there are known much stronger results and they are reviewed in chapter 20. Let us complement this by the following two remarks.

Remark 2.7. Ruzsa and Szemerédi (1978) related the proof $r_3(n) = o(n)$ (i.e., the upper bound) to a purely combinatorial problem.

Lemma 2.8. *Let* $(X, \mathcal{M})$ *be a triple system which contains on any set of* 6 *elements at most* 3 *triples. Then*

$$|\mathcal{M}| = o(|X|^2) .$$

From this Ruzsa and Szemerédi (1978) derived an $o(n)$ bound for $r_3(n)$ quite easily (see also Erdős et al. 1986). There is a similar combinatorial statement which is known to imply Szemerédi's theorem. However, this is still a conjecture at present.

Remark 2.9. The best lower bound for $r_3(n)$ is a classical result of Behrend (1946) and let us note that even the earlier weaker result of Salem and Spencer (1942), $r_3(n) > n^{1-\varepsilon}$ for a positive $\varepsilon > 0$ recently found a surprising application in a least expected area, namely in the fast multiplication of matrices.

Given two $n \times n$ matrices $A = (a_{ij})$, $B = (b_{ij})$, one computes their product

$$A \cdot B = (c_{ij})$$

by applying the definition $c_{ij} = \sum a_{ik} b_{kj}$ by means of $O(n^3)$ multiplications and additions. To a great surprise Strassen in his classical paper suggested a new method which consisted in breaking the matrices into blocks and recursively performing operations according to a complex pattern. The basis is his famous table for multiplying 2×2 matrices. This yielded a total number of operations $O(n^{2,7})$. Since then there have been numerous improvements of this result by Schönhage, Strassen, Coppersmith and Winograd. The best result so far was recently obtained by Coppersmith and Winograd (1990) who used the $O(n^{1-\varepsilon})$ lower bound for $r_3(n)$ to get an $O(n^{2,36})$ bound. The details are too complex to be presented here. Let us just remark that a dense sequence S without 3-term arithmetic progressions is used for a suitable indexing (hashing) of the very complex schema. Since one needs this for a fixed (very large) n, it does not matter whether one has a good algorithm for producing such a set S.

From the point of view of combinatorics, computer science has been a blessing. Nearly everything which has been studied over the years has found applications in some areas of computer science. In a sense, combinatorics has become a sort of "set theory" for computer science.

3. Some bounds

Perhaps the first question which one is tempted to consider is the problem of the actual size of a set which guarantees the validity of Ramsey's (and Ramsey type) theorem. One should try to resist this temptation since it is well known that Ramsey numbers are difficult to determine and even good asymptotic estimates are difficult to find (and improve).

We have already mentioned this in connection with the Van der Waerden numbers $W(n)$. For Ramsey numbers the situation is not as dramatic: It follows from Erdős and Rado (1950) and Erdős et al. (1965) that both upper and lower bound for $r(p, t, n)$ are of the form

$$t_p(c_{p,t} \cdot n) ,$$

where $t_p(x)$ is the two-variables version of the tower function $f_4 : t_p(x) = 2^{2^{\cdot^{\cdot^{\cdot^{2^x}}}}}$ (there are $p - 1$ 2's in the stack. See Duffus et al. (1995) for recent improvements.) However, for the most important case $t = 2$, the Ramsey number $r(p, 2, n)$ is bounded from below by the function t_{p-1} only. The question whether $r(3, 2, n)$ has an exponential lower bound belongs to the outstanding problems of the area. The estimates of Ramsey numbers form a rich spectrum of results. Below we state some of them.

Let us mention that this part of this chapter is expository since there are at least two recent surveys devoted to this area: the paper by Chung and Grinstead (1983) which concentrates on the exact results and the extensive paper by Graham and Rödl (1987) which concentrates mainly on the asymptotic bounds for most of the Ramsey-type results.

The progress has been rapid recently, and V. Rödl together with his coauthors systematically investigated asymptotic bounds for various Ramsey-related problems, see, e.g., Lefmann and Rödl (1993, 1995), Rödl and Rucinski (1993, 1995), Alon and Spencer (1992), Chen and Schelp (1993).

Recall that we denoted by $r(p, t, n)$ the minimum N which satisfies $N \rightarrow (n)^p_t$. Some special cases have a customary notation, e.g., $r(n) = r(2, 2, n)$ and $r(m, n)$ denotes the minimal number N for which every graph G with N vertices either satisfies $\omega(G) \geqslant m$ or $\alpha(G) \geqslant n$ (thus $r(n, n) = r(n)$; $r(m, n)$ is the corresponding off-diagonal number"). All known exact values of $r(m, n)$ are given in the following table:

m	3	3	3	3	3	3	3	4	4
n	3	4	5	6	7	8	9	4	5
$r(m,n)$	6	9	14	18	23	28	36	18	25

This is a great progress in this otherwise seemingly untractable problem: Since the first version of this paper two more numbers $r(4,5)$ and $r(3,8)$ were determined together with the first "hypergraph" Ramsey number $r(3,2,4)$. $r(3,8)=28$ is determined by McKay and Zhang (1992), $r(3,2,4)=13$ in McKay and Radziszowski (1991), $r(4,5)=25$ was announced by McKay and Radziszowski in March 1993. They started a massive computer-aided search and obtained strong improvements of bounds. For example for the much studied case $r(5)$ the current bounds are tight $43 \leq r(5) \leq 49$ (completed in June 1993). See Radziszowski (1993) for a list of recent results.

However, the problem of determination of exact values of $r(m,n)$ seems to be intractable in general. Perhaps motivated by this Chvátal and Harary in a series of papers (see, e.g., Chvátal and Harary 1972) suggested studying "generalized Ramsey numbers" defined as follows: $r(G,H)$ is the minimum N with the following property:

If the edges of K_N are colored by blue and red then either there exists a blue subgraph isomorphic to G or a red subgraph isomorphic to H. Also set $r(G)=r(G,G)$. Thus $r(K_m,K_n)=r(m,n)$. However, the number $r(G,H)$ seems to be easier to determine especially if one of the graphs G and H is a sparse graph. Let us list some particularly elegant examples:

(1) $r(T_m,K_n)=(m-1)(n-1)+1$ (Chvátal 1977) (here T_m is a fixed tree with m vertices).

(2) $r(n \cdot K_3)=5 \cdot n$ for $n \geq 2$ (Burr et al. 1975) (here $n \cdot K_3$ is the disjoint union of n triangles; of course for $n=1$ we have a different formula).

(3) $r(G_n) \geq \lfloor (4n-1)/3 \rfloor$ for every connected graph with n vertices and for every $n \geq 3$ there are graphs which prove that this inequality is sharp (Burr and Erdős 1976).

These results were generalized and a great variety of related results was obtained. See, e.g., Burr and Erdős (1975), Erdős et al. (1985, 1989, 1995), Faudree and Simonorits (1992), Erdős and Faudree (1992), Chvátal et al. (1983), Sidorenko (1991, 1994). We want to mention the following two striking results (both relying heavily on the Szemerédi's (1976) regularity lemma.

Theorem (Chvátal et al. 1983). *For every k there exists constant c_k such that $r(G) \leq c_k \cdot n$ for every graph G with n vertices and maximal degree $\leq k$.*

Theorem (Chen and Schelp 1993). *There exists a constant c such that $r(G) \leq c \cdot n$ for every planar graph G with n vertices.*

Very recently Rödl and Thomas (1995) extended this result to graphs not containing a subdivision of K_k.

These results are partial results towards a solution of the following conjecture of Burr and Erdős (1975).

Conjecture. For every constant k there exists a constant $c(k)$ such that $r(G) \leq c(k) \cdot n$ for every graph G with n vertices and at most $k \cdot n$ edges.

Let us mention another recent striking conjecture due to Loebl:

Conjecture. $r(T_n) \leq 2n$ for every tree T with n vertices.

If one considers the distribution of papers within Ramsey theory then probably it is fair to guess that the number of papers devoted to generalized Ramsey numbers has the highest frequency. It is only fitting that Burr, Faudree and Schelp are planning to prepare a monograph devoted to this subject.

Let us return to the classical Ramsey numbers. We shall list a sample of asymptotic results.

3.1. *Upper asymptotics*

Generally, there is a big difference between proof of lower and upper bounds for Ramsey numbers. The upper bounds yield a proof of Ramsey's theorem itself. There are not many variations here and consequently all upper bounds are inductive proofs. (Particularly there is no probabilistic upper bound.) Let us be more specific.

The classical Erdős and Szekeres (1935) bound

$$r(m+1, n+1) \leq \binom{m+n}{m}$$

is a consequence of the recursion

$$\begin{aligned} r(m+1, n+1) &\leq r(m, n+1) + r(m+1, n) \\ &= ((r(m, n+1) - 1) + (r(m+1, n) - 1) + 1) + 1 \end{aligned}$$

and the induction step is indicated in fig. 3.1.

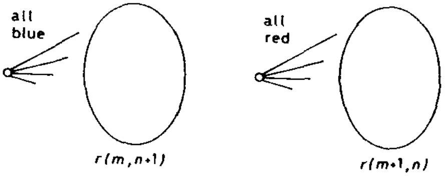


Figure 3.1.

The Erdős–Szekeres bound has been improved several times but only after fifty years was it shown that $r(m+1, n+1) = \mathrm{O}((\binom{m+n}{n}))$, which was proved independently by Rödl and Thomason. Their bounds are the following:

(1) $r(m+1, n+1) \leq c\binom{m+n}{m}/(\log(m+n))^{c'}$ for positive constants c and c'

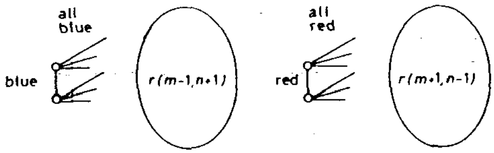


Figure 3.2.

(Rödl 1986a). See Graham and Rödl (1987) for the following weaker result:

$$r(m+1, n+1) \leqslant 6\binom{m+n}{m}/\log\log(m+n) .$$

(2) $r(n+1, n+1) \leqslant \binom{2n}{n}/\sqrt{n}$ (Thomason 1987, 1988).

The basis of both proofs is similar in that they systematically use the following recursion (Walker 1971).

We get either a blue K_m or a red K_n if either there exists a blue edge contained in $r(m-1, n+1)$ blue triangles or a red edge contained in $r(m+1, n-1)$ red triangles. See fig. 3.2.

This approach is convenient since the number of monochromatic triangles is estimated very accurately by the following result of Goodman (1959) which is interesting in its own right.

Theorem 3.1. *Denote by $k_3(G)$ the number of triangles which are contained in the graph G. Let G be a graph with n vertices and $m = \rho\binom{n}{2}$ edges. Then*

$$c_3(G) + c_3(\bar{G}) \geqslant (\rho^3 + (1-\rho)^3)\binom{n}{3} - \rho(1-\rho)\binom{n}{2} , \tag{3.1}$$

where $\bar{G}$ is the complement of the graph G.

Proof. Let $d_1, \ldots, d_n$ be the degrees of the vertices of G. We have $\sum d_i = \rho\binom{n}{2}$ and also

$$\sum d_i^2 \geqslant \sum d_i \cdot \frac{\sum d_i}{n} .$$

Observe that $d_i \cdot (n-1-d_i)$ is the number of triples containing a given vertex x of degree d_i which fail to form a triangle in both G and $\bar{G}$ regardless of whether the edge not containing x belongs to G or $\bar{G}$. Summarizing, we get

$$\begin{aligned} c_3(G) + c_3(\bar{G}) &\geqslant \binom{n}{3} - \tfrac{1}{2}\sum_i d_i(n-1-d_i) \\ &= \binom{n}{3} - \frac{n-1}{2}\sum d_i + \tfrac{1}{2}\sum d_i^2 \\ &\geqslant \tfrac{1}{6}[n(n-1)(n-2) - 3pn(n-1)^2 + 3p^2n(n-1)^2] \end{aligned}$$

which gives (3.1). □

(3.1) implies that every graph G with n vertices satisfies

$$c_3(G) + c_3(\bar{G}) \geq \tfrac{1}{4}\binom{n}{3} + \mathrm{O}\left(\binom{n}{2}\right).$$

This is the expected estimate which we obtain for $c_3(G)$ of the random graph G_n (from 8 graphs on 3 points we are interested in exactly two of them).

More generally, for a fixed $l \geq 2$, denote by $c_l(G)$ the number of complete subgraphs of order l in G and put

$$c_l(n) = \min\{(c_l(G) + c_l(\bar{G}))/\tbinom{n}{l}\},$$

where the minimum is taken over all graphs G with n vertices.

Ramsey's theorem implies that $c_l(n) > 0$ providing n is sufficiently large. One can also easily observe that $c_l(n)$ is an increasing function and thus let c_l denote the $\lim_{n\to\infty} c_l(n)$. Erdős (1962) conjectured that $c_l = 2/2^{\binom{l}{2}}$ for every $l \geq 2$.

This is true for $l = 2$ (trivially) and for $l = 3$ by virtue of the above result of Goodman.

Let us remark that this conjecture reflects again the value expected from a random coloring of K_n. However, the Erdős conjecture has been disproved for every $l > 3$ by Thomason (1987).

Let us remark that some off-diagonal numbers have been estimated rather efficiently. Perhaps the most important result is the following bound due to Shearer (1983) who improved and exploited a fundamental method of Ajtai et al. (1980).

Theorem 3.2.

$$r(n, 3) < \frac{n^2}{\log(n/e)}.$$

This bound and the method of its proof has found many applications (see, e.g., Ajtai et al. 1981a,b and Komlós et al. 1981).

3.2. Lower asymptotics

Lower bounds use different methods.

In proving $r(n) > N$ we have to exhibit a graph G with N vertices (and $N \gg n$) such that neither G nor its complement contains K_n. The search for G is difficult and the best results are usually obtained by probabilistic methods. In fact it is possible to say that the problem of estimating lower bounds for Ramsey numbers has been the cradle of the probabilistic method in combinatorics. Since this is covered in chapter 33, we only add a few remarks. The best available exponential lower bound for $r(n)$ is

$$r(n) > (1 + \mathrm{O}(1))\frac{\sqrt{2}}{\mathrm{e}} n \cdot 2^{n/2}.$$

However, the search for an explicit graph of size (say) $2^{n/2}$ which would demonstrate this lower bound has been so far unsuccessful. This is not an entirely satisfactory situation since it is believed that such graphs share many properties with random graphs and thus they could be good candidates for various lower bounds, for example, in theoretical computer science for lower bounds for various measures of complexity. (See the papers Chung et al. 1989 and Thomason 1987 which discuss properties of pseudo- and quasi-random graphs also see chapter 6.) This use of Ramsey's theorem (and Ramsey-type statements) is documented, e.g., by papers of Vilfan (1976), Pudlák (1984), Alon (1986). An interesting application of Ramsey's theorem is in a paper of Grinčuk (1989) where he uses the *infinite* Ramsey theorem to prove a nonlinear lower bound to the size of switching circuits computing symmetric Boolean functions.

The quasi-random property of graphs involved in the lower bounds for Ramsey numbers was studied in the broader context by a number of authors (Erdős and Hajnal 1977, Nešetřil and Rödl 1979b, Rödl 1986b, Thomason 1987, Chung et al. 1989, Chung and Graham 1991). This may be viewed also as a part of the recent trend to *derandomize* (see Alon and Spencer 1992) the non-constructive proofs. As mentioned earlier, apart from the classical counting techniques mostly due to P. Erdős, in Ramsey theory other useful tools involve Lovász' local lemma (Erdős and Lovász 1975) and Szemerédi's regularity lemma (Szemerédi 1976) both of nonconstructive nature. These results recently found a constructive version (Alon et al. 1994, Beck 1991) see also an ad hoc argument in a Ramsey geometrical setting (Alon et al. 1995).

The best constructive lower bound for Ramsey numbers $r(n)$ is due to Frankl and Wilson. This improves on an earlier construction of Frankl (1977) who found a first constructive superpolynomial lower bound.

The construction of Frankl–Wilson graphs is simple:

Let p be a prime number, put $q = p^3$. Define the graph $G_p = (V, E)$ as follows:

$$V = \binom{[q]}{p^2 - 1} = \{F \subseteq \{1, \ldots, p^3\} : |F| = p^2 - 1\} ,$$

$$\{F, F'\} \in E \text{ iff } |F \cap F'| \equiv -1 \pmod{q} .$$

The graph G_p has $\binom{p^3}{p^2-1}$ vertices. However, the Ramsey properties of the graph G_p are not trivial to prove: It follows only from deep extremal set theory results due to Ray-Chaudhuri and Wilson (1975) and Frankl and Wilson (1981) that neither G_p nor its complement contain K_n for $n \geqslant \binom{p^3}{p-1}$. (Compare also a recent article by Frankl 1990 and chapter 24.)

Let us finally remark that until very recently the best lower bound for $r(n, 3)$ was $O(n^2/(\log n^2))$ (Spencer 1977). Somewhat surprisingly Kim (1995) proved that the upper bound in Theorem 3.2 is tight up to a constant factor:

Theorem 3.3.

$$r(n, 3) \geqslant c \frac{n^2}{\log n} .$$

$r(n,3)$ is the only non-trivial infinite family of (classical) Ramsey numbers with known asymptotics.

The best constructive lower bound for Ramsey numbers $r(3,n)$ is of order $n^{3/2}$. This follows from a recent elegant construction given by Alon (1994), see also Chung et al. (1993).

3.3. *Even larger numbers*

The bounds on Ramsey numbers indicate that Ramsey's theorem is a rather "non-effective" statement which involves a study of very large sets. In fact, this non-effective character of Ramsey's theorem has been known for a long time and especially the following infinite Ramsey theorem (Ramsey 1930) has been studied from this point of view.

Infinite Ramsey theorem 3.4. *For all positive integers t and p and every partition of the set $\binom{X}{p}$ of all p-element subsets of an infinite set X into t classes, one of the classes contains all p-element subsets of an infinite set Y.*

Symbolically we write $\omega \to (\omega)^p_t$, ω standing for the order type of the set $\mathbb{N}$ of all natural numbers.

Proof. The proof of the Infinite Ramsey theorem is similar to the proof of the finite Ramsey theorem and, in fact, conceptually simpler:

We proceed by induction on p. The case $p=1$ is the infinite pigeonhole principle: One of the classes of a finite partition of an infinite set has to be infinite.

In the induction step let $c: \binom{\mathbb{N}}{p+1} \to [t]$ be a given coloring. Define a coloring $c_0: \binom{\mathbb{N}-\{0\}}{p} \to [t]$ by $c_0(A) = c(A \cup \{0\})$. Using the induction hypothesis there exists an infinite set $X_0 \subseteq \mathbb{N}$ such that $c_0 | \binom{X_0}{p}$ is a constant mapping; let i_0 denote the constant. Set $X'_0 = X_0 - \{\min X_0\}$, set $x_1 = \min X_0$, and define the coloring

$$c_1: \binom{X'_0}{p} \to [t] \quad \text{by } c_1(A) = c(A \cup \{x_1\}).$$

There exists infinite $X_1 \subseteq X'_0$ such that $c_1 | \binom{X_1}{p}$ is a constant denoted by i_1. Continuing this way we construct elements $x_0 = 0 < x_1 < x_2 < \cdots$ and numbers i_0, i_1, $i_2, \ldots$ Define the coloring $c': \{x_0, x_1, x_2, \ldots\} \to [t]$ by

$$c'(x_j) = i_j .$$

Find infinite Y such that $c'|_Y$ is a constant mapping and check that Y is a c-homogeneous set. □

Only recently has the impact of the Infinite Ramsey theorem on finite combinatorics been recognized. Let us be more precise and let us briefly describe this interesting recent development on the border line of combinatorics and logic.

The theory of finite sets is related to the most frequent axiomatization of the natural numbers known as Peano arithmetic (PA).

The Peano axioms consist of the following formulas which express the basic properties of arithmetic:

(i) $\forall x(x+\underline{0}=x)$

(ii) $\forall x \forall y(x+(y+\underline{1})=(x+y)+\underline{1})$

(iii) $\forall x(x \cdot \underline{0})$

(iv) $\forall x \forall y(x \cdot (y+\underline{1})=(x \cdot y)+x)$

(v) $\forall x \forall y(x<y \leftrightarrow \exists z((x+z)+\underline{1}=y)$

(vi) For any arithmetic formula $\Theta(x, y)$, it holds that

$$\forall x((\Theta(x,\underline{0}) \wedge \forall y(\Theta(x, y) \rightarrow \Theta(x, y+\underline{1}))) \rightarrow \forall y \Theta(x, y)) .$$

(Here, a formula is arithmetic if it is built up from binary predicates $+,\cdot,<$ and constant symbols $\underline{0}$, $\underline{1}$ by logical connectives and quantifiers.)

These axioms express properties of natural numbers and many (in fact, at the time of the creation of the system it was believed all) true statements about natural numbers seem to be possible to be deduced from them.

Strictly speaking the formulas about natural numbers do not seem to capture properties of finite sets. However, one can encode finite sets by natural numbers and the finite set theory may be transformed into arithmetic. It appears that PA is equivalent to the theory of finite sets. This means "usual" set theory together with the negation of the axiom of infinity. (This axiom may be expressed as follows: There is no set X which satisfies $x \in X \Rightarrow \{x\} \in X$.)

Most of the finite combinatorics is carried out within the theory of finite sets. But not all, as is illustrated by the following.

Theorem 3.5 (Paris and Harrington 1977). (1) *Let p, t, n be positive integers. Then there exists an N such that $N \underset{*}{\rightarrow} (n)^p_t$. Here the (modified) partition arrow $\underset{*}{\rightarrow}$ has the following meaning: For every partition $\binom{[N]}{p}$ into t classes there exists a set $X \subseteq [N]$ with the following properties*:

(i) *$\binom{Y}{p}$ belongs to one of the classes of the partition*;

(ii) *$|Y| \geq n$*;

(iii) *$|Y| \geq \min Y$.*

(2) *The statement* (1) *cannot be proved within the theory of finite sets* (*or deduced from the Peano axioms*).

Proof of (1). Suppose that (1) fails to be true for a particular choice of p, t, n. This means that for every N there exists a "bad" coloring $c_N : \binom{[N]}{p} \rightarrow [t]$ such that at least one of the conditions (i), (ii) and (iii) is violated. Observe that a restriction of a bad coloring to a subset is again a bad coloring. Thus for every N we may find a bad coloring c^N such that c^i is a restriction of c^j for $i<j$. Denote by c the coloring of $\binom{\mathbb{N}}{p}$ which extends all the colorings c^i. Now invoke the Infinite Ramsey theorem: There exists an infinite set $X=\{x_1<x_2<\cdots\}$ such that all p-element subsets of X are monochromatic under c. Put $r=\max\{x_1, n\}$. Then the set $Y=\{x_1, \ldots, x_r\}$ satisfies conditions (1), (2) and (3). □

Of course, (2) is the important part of the theorem of Paris and Harrington. Their original proof was motivated by a careful study of models of Peano arithmetic.

More combinatorial insight was gained by a different approach due to Ketonen and Solovay (1981). They proceeded as follows:

First, define a hierarchy of rapidly growing functions $f_\alpha : \omega \to \omega$, $\alpha \leq \varepsilon_0$. Here α is an ordinal number and ε_0 is the first ordinal number λ which satisfies $\omega^\lambda = \lambda$; thus ε_0 is the tower of ω's of "height ω":

$$\omega^{\omega^{\omega^{\cdot^{\cdot^{\cdot}}}}} .$$

The functions f_α are defined by transfinite induction on α:

$$f_0(n) = n + 1 ,$$
$$f_1(n) = 2 \cdot n ,$$
$$f_{\alpha+1}(n) = \underbrace{f_\alpha \circ \cdots \circ f_\alpha}_{n}(1) ;$$

for a limit ordinal α we put

$$f_\alpha(n) = f_{\alpha_n}(n) .$$

Here (for a limit ordinal number α) $\{\alpha_n\}$ is the sequence with the *quickest* convergence to α. It is here where one uses properties of ordinals $\alpha < \varepsilon_0$: Every such ordinal can be written uniquely in the Cantor normal form

$$\alpha = \omega^{\alpha_1} k_1 + \omega^{\alpha_2} k_2 + \cdots + \omega^{\alpha_r} k_r ,$$

where $\alpha_1 > \cdots > \alpha_n \geq 0$ are ordinal numbers strictly less than α and $k_1, \ldots, k_n$ are positive integers. α_λ is then defined by means of the last term $\omega^{\alpha_r} k_r$. For example for $\alpha = \omega^{\omega^2} + \omega^{\omega+3} 5$ we define

$$\alpha_n = \omega^{\omega^2} + \omega^{\omega+3} 4 + \omega^{\omega+2} n$$

and for $\alpha = \omega^{\omega^2} + \omega^{3\omega} 5$ we define

$$\alpha_n = \omega^{\omega^2} + \omega^{2\omega+n} 5 .$$

Besides of these elements of ordinal arithmetic we shall make use of only one result of mathematical logic.

Theorem 3.6 (Wainer 1970). *If $f : \omega \to \omega$ is recursive and provably total (in PA), then there exists $\alpha < \varepsilon_0$ and a positive integer n_0 such that $f(n) < f_\alpha(n)$ for all $n > n_0$ (i.e., f is eventually dominated by a function from the hierarchy).*

Now finally consider the function $r^*(p, t, n)$ defined as follows:

$$r^*(p, t, n) = \min\{N : N \underset{\cdot}{\to} (n)_t^p\} .$$

It follows from the first part of the Paris–Harrington theorem that $r^*(p, t, n)$ is a total function. Ketonen and Solovay (1981) now prove by induction on α that the diagonal function

$$r^*(p, p, p+1)$$

eventually dominates every function f_α, $\alpha < \varepsilon_0$. Combining this with the Wainer theorem, this shows that $r^*(p, p, p+1)$ fails to be provably total recursive.

The Ketonen–Solovay approach is very interesting from a combinatorial point of view since it provides a combinatorial reason *why* the Paris–Harrington theorem fails to be provable: the Ramsey function $r^*(p, t, n)$ grows so fast that one cannot express it by standard operations.

In other words, the standard combinatorial approach to a problem "estimate numbers $r^*(p, t, n)$" does not always have in general a solution – at least in the sense in which it has been asked.

This is a new phenomenon in our old finite (safe and paradox-free) combinatorics.

This is not to say that one cannot obtain interesting results in special cases. Indeed, this has been done for numbers $r^*(n) = r^*(2, 2, n)$ by Erdős and Mills (1981), where, in particular, it was proved that

$$n^{2^{\alpha n}} < r^*(n) < n^{2^{\beta n}}$$

for convenient constants α and β.

Recently, Loebl and Nešetřil (1992) found a different and simple proof of the Ketonen–Solovay result. This has been found in a broader context and the author cannot resist mentioning one more result which follows the program of the combinatorial study of undecidability. (See Nešetřil and Thomas 1987, Paris 1990, and Loebl and Nešetřil 1991 for surveys of this development.)

One such example may be derived from Kruskal's theorem (see chapter 5 of this book): Let us denote by KT(c) the following statement: "*For every $k \geq 0$ there exists $n_c(k)$ with the following property: If $T_1, \ldots, T_{n(k)}$ are trees which satisfy $|T_i| \leq k + c\log(i)$ for all $i = 1, \ldots, n_c(k)$, then there are two distinct indices i and j such that T_i contains a subtree homeomorphic to T_j.*"

Now we have a result obtained by Loebl and Matoušek (1987).

Theorem 3.7.

(i) *The statement* KT(c) *is true for every c.*

(ii) *The statement* KT(1/2) *is provable in the theory of finite sets.*

(iii) *The statement* KT(2) *fails to be provable in the theory of finite sets.*

Sketch of Proof. (i) In fact much more is true. If $f : \mathbb{N} \to \mathbb{N}$ is any function then there exists $n(f)$ such that the following holds: If $T_1, \ldots, T_{n(f)}$ are trees satisfying $|T_i| \leq f(i)$ then there are two distinct values i and j such that T_i contains a subtree homeomorphic to T_j.

Suppose that this fails to be true for a particular function f. Thus, for every n, there exists a family $T_1^n, \dots, T_n^n$ of trees which violates (i). But since there are only finitely many possibilities for T_1^n we find an infinite set I_1, such that $T_1^n = T_1$ for all $n \in I_1$. Then we find an infinite set $I_2 \subseteq I_1$ such that $T_2^n = T_2$ for all $n \in I_2$ etc. For the resulting sequence $T_1, T_2, \dots, T_n, \dots$ we apply Kruskal's theorem and obtain two trees T_i and T_j one of which contains a homeomorphic copy of the other. However, $T_i = T_i^n$, $T_j = T_j^n$, for some n, which is a contradiction.

(ii) Follows from the fact that there are $< 4^n$ pairwise non-isomorphic trees with n vertices.

(iii) is based on the analysis of the so-called Hercules–Hydra game which was analysed in the papers Kirby and Paris (1982), Loebl (1988, 1992) and Nešetřil (1984). □

Theorem 3.7. strengthens an earlier result of Friedman (see Simpson 1985) and it is perhaps the most striking current combinatorial example of unprovability.

Let us end this chapter by the following remark: However surprising it may be at first glance, it is only fitting to mention in a chapter on Ramsey theorem a result related to well-quasi-ordering theory (WQO). Both theories share many similarities (see, e.g., Leeb 1973) and several results have proved to be mutually fruitful. For example Nash-Williams proved the following strengthening of Ramsey's theorem in the context of WQO-theory.

Theorem 3.8 (Nash-Williams 1965). *Let $\mathcal{M}$ be a Sperner-system of subsets of an infinite set X (i.e., no two distinct elements of $\mathcal{M}$ are inclusion-related). Then for every finite partition $\mathcal{M} = \mathcal{M}_1 \cup \cdots \cup \mathcal{M}_t$ there exists an infinite subset Y of X such that all members of $\mathcal{M}$ which are subsets of Y belong to one of the classes of the partition.*

(This generalizes the Infinite Ramsey theorem; $\binom{X}{p}$ is a Sperner system.)

It is interesting to note (as observed already by Erdős and Rado) that Ramsey's theorem does not hold for partitions of infinite subsets in a very strict sense.

Theorem 3.9 (Erdős, Rado). *For every infinite set X of cardinality K holds*

$$K \not\to (\omega)_2^\omega .$$

Sketch of a proof (Nešetřil and Rödl 1985). For any well ordered set X define a graph $G = (V, E)$ as follows:

$$V = \binom{X}{\omega}, \qquad \{A, B\} \in E \text{ iff } A = \{a_0 < a_1 < \cdots\} .$$

The graph does not contain a circuit thus $\chi(G) \leq 2$ which in turn induces a 2-coloring of $\binom{X}{\omega}$ violating $K \to (\omega)_2^2$. □

Although a simple argument, yet this involves the axiom of choice. This *nonconstructive* feature of Ramsey-type theorems has been studied in great detail in the context of mathematical logic (Jockush 1972). Also theorems 3.5 and 3.7 can be interpreted in this framework. In another important direction many results were obtained about the validity of $\omega \rightarrow (\omega)_2^\omega$ for *constructive partitions*. The constructive partitions are usually defined in topological terms as open, Borel or Baire-partitions. Theorem 3.8 played an important role in this development of "topological Ramsey theorems" by Galvin, Prikry and others. We shall comment on this later in section 4.4. See the survey article by Carlson and Simpson (1990).

This development has a finite analogy: if we restrict partitions by a structural condition it is expected that a (much) larger homogeneous set will be found. Various types of such restrictions are provided by the following papers: Alon (1990) (solving a conjecture of Babai-partitions defined by polynomials); Erdős et al. (1983), Babai (1985), Alon et al. (1991) (various local conditions such as "anti-Ramsey" theorems: no intersecting edges get the same color); Sparks (1993) (definable colorings); Larman et al. (1992) (graphs defined by the intersection of n convex sets have either ω or $\alpha \geqslant n^{1/5}$). The paper by Erdős and Hajnal (1977) is one of the papers who initiated the development of quasi-random graphs and it contains the following problem which may be viewed as a root question for this kind of problem.

Problem (Erdős and Hajnal 1977). Suppose H is a fixed graph and a graph G with n vertices contains no induced subgraph isomorphic to H. Is it then necessarily true that either $\alpha(G) \geqslant n^\varepsilon$ or $\omega(G) \geqslant n^\varepsilon$ for some fixed $\varepsilon > 0$?

Finally, returning to our main theme, let us remark that recently Kříž and Thomas (1990) have applied the ordinal-type techniques (known in WQO) to countable Ramsey theory.

4. Full Ramsey-type theorems

Ramsey's and Van der Waerden's theorems together with the companion results of Schur and Hales and Jewett are milestones of Ramsey theory. We call them *full Ramsey-type theorems* as they assert that *every sufficiently large structure* for every partition of a given type contains a given homogeneous structure (graph, cube, segment of integers, etc.).

In others words if we have a "bad partition" of a structure S then there is an absolute bound on size of S.

Natural examples of full Ramsey-type theorems are difficult to find and the list seems to be finite. With a few variations these are the theorems listed in this section.

4.1. *Rado sets of linear equations*

Both Schur's theorem and Van der Waerden's theorem fit the following more general schema: Let $A = (a_{ij})$ be an integer m by n matrix. Then for every $t \geqslant 1$, there exists an integer $N = N(A, t)$ which has the following properties: If the set

$\{1, 2, \dots, N\}$ is partitioned into t classes then in one class of the partition there is a solution of a system of equations:

$$\begin{aligned} a_{11}x_1 + \cdots + a_1 x_n &= 0 \\ a_{21}x_1 + \cdots + a_{2n}x_n &= 0 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= 0 \end{aligned} \tag{4.1}$$

For the Schur theorem consider the matrix A consisting of the single row

$$A = (1, 1, -1)\,.$$

For the Van der Waerden theorem consider, e.g., the following matrix:

$$A = \begin{pmatrix} 1 & -1 & & & & & \\ & 1 & -1 & & & & \\ & & 1 & -1 & & & \\ & & & 1 & -1 & & \\ & & & & \ddots & & \\ & & & & & 1 & -1 \\ & & & & & & 1 \end{pmatrix}$$

which corresponds to a stronger statement (also conjectured by Schur, see the introduction of Schur 1973) and proved by his student Brauer (1928) that in arbitrary finite partition of integers we can find in one of the classes the arithmetic progression a_0, $a_0 + d, \dots, a_0 + (n-1)d$ together with the difference d.

We can abbreviate (4.1) by writing

$$Ax = 0\,, \quad x = (x_1, \dots, x_n)^{\mathrm{T}}$$

and even more considerably we can denote the system (4.1) of equations by $\mathcal{S}(A)$.

The basic (and at first glance perhaps too ambitious) problem is to characterize those integral matrices A for which a result analogous to Schur's and Van der Waerden's theorems holds. This leads to the following notions:

The set of equations $Ax = 0$ is said to be *partition regular* if for any finite partition of positive integers N there is always a solution of the system (4.1) in one of the classes. (By compactness, it does not matter whether we formulate the finite or infinite version.)

Note that obviously not every set of equations is partition regular, e.g., consider $x = 2y - 1$ and a parity argument. However, one should say that it is surprising that one can characterize all partition regular systems as follows:

An m by n matrix $A = (a_{ij})$ is said to satisfy the *columns condition* if it is possible to order its column vectors $a_1, \dots, a_n$ so that for some choice of indices $1 \leq n_1 < n_2 < \cdots < n_t = n$, if we set

$$b_i = \sum_{j=n_{i-1}+1}^{n_i} a_j$$

then

(i) $b_1 = 0$,

(ii) for $1 < i \leq t$, the vector b_i can be expressed as a *rational* linear combination of columns a_j, $1 \leq j \leq n_{i-1}$.

Now we can formulate the following.

Theorem 4.1 (Rado 1933). *The system $Ax = 0$ is partition regular if and only if A satisfies the columns condition.*

In neither direction this is a trivial result. For modern write-up see Graham et al. (1980).

Perhaps it is interesting to note that in one direction (necessity) it is sufficient to use partitions of integers modulo a prime. This is the easier part. In order to prove sufficiency one generalizes tricks which are involved in a proof of the Hales–Jewett theorem. However, most fitting in this context is to use Deuber's axiomatization of subsets of positive integers for which the analogue of Rado's theorem holds. More precisely: a set X of positive integer is called *large*, if for any partition regular system $Ax = 0$ and any finite partition of X, one of the classes of partition contains a solution of $Ax = 0$.

Using this terminology Rado's theorem reduces to saying that the set of positive integers is large. The following proved a long-standing conjecture of Rado (1933).

Theorem 4.2 (Deuber 1973). *If X is a large set then for every finite partition of X, one of the classes is again a large set.*

In order to obtain this result, Deuber provided the axiomatization of large sets by means of the following concept:

A set X of positive integers is called an *(m, p, c)-set* if there are positive integers $y_1, \ldots, y_m$ such that X is the set of all linear combinations of the form

$$\sum_{i=1}^{m} \lambda_i y_i ,$$

where the vector $(\lambda_1, \ldots, \lambda_m)$ itself has form (for some $i < m$)

$$(0, 0, \ldots, c, \lambda_{i+1}, \ldots, \lambda_m)$$

and $|\lambda_j| \leq m$, $j = i + 1, \ldots, m$.

It has been shown by Deuber (1973) (see Leeb 1973 for a simplified proof) that for every choice of positive integers m, p, c and t there exist M, P, C such that any t-coloring of an (M, P, C)-set always contains a monochromatic (m, p, c)-set.

This implies Rado's theorem and Deuber's theorem since regular sets may be characterized by (m, p, c)-sets: a set X is large if it contains an *(m, p, c)-set* for every choice of positive integers (m, p, c).

We do not prove Theorem 4.2 here, see, e.g., Graham et al. (1980). Let us indicate at least how the notion of an *(m, p, c)-set* naturally emerges: We prove that a regular system of equations always has solutions which form an (m, p, c)-*set*.

This is a consequence of the columns condition. Let us be more specific. Let A be an m by n matrix satisfying the columns condition. Let $1 \leq n_1 < n_2 < \cdots < n_l = n$ be the corresponding partition of columns of A (denoted by $a_1, \ldots, a_n$). Since we are interested in solutions of a system of equations $Ax^{\mathrm{T}} = 0$, we may assume that the rows of A are linearly independent. Using the columns condition, a set of $n - m$ linearly independent solutions of $Ax^{\mathrm{T}} = 0$ has the following form:

$$\begin{array}{l}
(\overbrace{1 \quad ,\ldots, \quad 1}^{n_1},\overbrace{0,\ldots,0}^{n_2},0,\ldots,0,\overbrace{0,\ldots, \quad 0}^{n_l})^{\mathrm{T}}\\
(\kappa_{2,1} \quad ,\ldots, \quad \kappa_{2,n_1}, 1,\ldots,1,0,\ldots,0,0,\ldots, \quad 0)^{\mathrm{T}}\\
\qquad \vdots\\
(\kappa_{l,1} \quad ,\ldots, \quad \kappa_{l,n_1}, \quad ,\ldots, \quad ,\ldots, \quad 1,\ldots, \quad 1)^{\mathrm{T}}\\
(\kappa_{l+1,1} \quad ,\ldots, \quad \kappa_{l+1,n_1}, \quad ,\ldots, \quad ,\ldots, \quad \ldots, \kappa_{l+1,n})^{\mathrm{T}}\\
\qquad \vdots\\
(\kappa_{n-m,1},\ldots,\kappa_{n-m,n_1}, \quad ,\ldots, \quad ,\ldots, \quad \ldots,\kappa_{n-m,n})^{\mathrm{T}}
\end{array}$$

with rational entries. Thus we may multiply by a sufficiently large integer, say c, to obtain linearly independent integer vectors of the form

$$\begin{aligned}
x_1 &= (c, \ldots, c, 0, \ldots, 0, \ldots, 0, \ldots, 0)^{\mathrm{T}}\\
x_2 &= (\lambda_{2,1}, \ldots, \lambda_{2,n_1}, c, \ldots, c, \ldots, 0, \ldots, 0)^{\mathrm{T}}\\
&\vdots\\
x_l &= (\lambda_{l,1}, \ldots, c, \ldots, c)^{\mathrm{T}}\\
&\vdots\\
x_{n-m} &= (\lambda_{n-m,1}, \ldots, \lambda_{n-m,n_1}, \ldots, \lambda_{n-m,n})^{\mathrm{T}}\,.
\end{aligned}$$

Set $p = |\max \lambda_{ij}|$. Now let x be a solution of $Ax = 0$. Then there are $y_1, \ldots, y_{n-l}$ such that

$$x = \sum_{i=1}^{n-m} y_i x_i$$

and thus each entry of x belongs to an $(n - m, p, c)$-set. □

The above results were generalized in several directions some of which we shall describe below. Other generalizations are covered by Deuber (1975b), Voigt (1980), Bergelson et al. (1991, 1995) (non-homogeneous systems of equations, infinite systems, equations in general algebraic structures).

4.2 *Parameter sets (the Graham–Rothschild theorem)*

The partition theorem for parameter sets generalizes the Hales–Jewett theorem in the same way that Ramseys' theorem generalizes the pigeonhole principle. Parameter sets are higher-dimensional analogues of combinatorial lines. We state the theorem after introducing the necessary notions.

Throughout this section let $A = \{1,2, \ldots, a\}$ be fixed. We call A the *alphabet*.

Consider the N-dimensional cube A^N over A. An *n-parameter set* in A^N is a set $\boldsymbol{A}$ of the following form: there are non-empty disjoint subsets $A_1, A_2, \ldots, A_n$ of $\{1, 2, \ldots, N\}$ and a point $f^0 = (f_1^0, \ldots, f_N^0)$ of A^N such that $\boldsymbol{A}$ is the set of all points $f \in A^N$ which satisfy

$$f_i = \begin{cases} f_i^0 & \text{for } i \not\in \bigcup\limits_{i=1}^{n} A_j\,, \\ f_{i'} & \text{for } i, i' \in A_j,\ j = 1, \ldots, n\,. \end{cases}$$

We also say that the parameter set $\boldsymbol{A}$ has been determined by $(f^0, A_1, \ldots, A_n)$. We always choose the order of $A_1, \ldots, A_p$ such that

$$\min A_1 < \min A_2 < \cdots < \min A_n\,.$$

A_i is called the ith *moving coordinate*. It is clear that $|\boldsymbol{A}| = a^n$ and we can schematically visualize $\boldsymbol{A}$ by a picture like fig. 4.1.

A^N itself is an N-parameter set. If $\boldsymbol{A}$ is an n-parameter set and $p \leq n$ denote by $\binom{\boldsymbol{A}}{p}$ the set of all p-parameter subsets of $\boldsymbol{A}$.

It is fair to say that one of the turning points in Ramsey theory was the following theorem established by Graham and Rothschild (1971).

Theorem 4.3 (Ramsey's theorem for parameter sets). *Let n, t be positive integers, let p be a nonnegative integer, and let A be a finite alphabet. Then there exists $N = \mathrm{GR}(A, p, t, n)$ with the following property: If C is an N-parameter set and $\binom{C}{p}$ is colored by t-colors then there exists an n-parameter set $\boldsymbol{B}'$ for which $\binom{\boldsymbol{B}'}{p}$ is a monochromatic set.*

We denote the validity of this statement again by $N \rightarrow (n)_t^p$ having in mind that this arrow is shorthand notation which should be interpreted for parameter sets. The proof of the Ramsey property of parameter sets was originally rather difficult (Graham and Rothschild 1971); see Leeb (1973) and Deuber and Voigt (1982) for simplified proofs. We give here a proof which proceeds by complete analogy with Proof II of Ramsey's theorem stated in section 1.

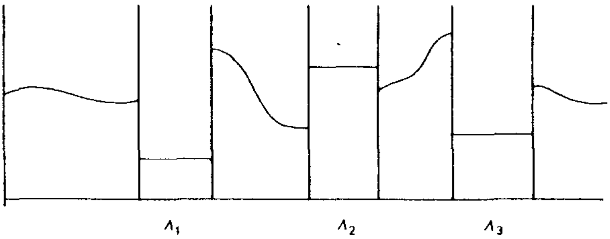


Figure 4.1.

Proof. The beginning of a p-parameter set $\boldsymbol{A}$ which is determined by $(f^0, A_1, \dots, A_p)$ is the function $f^0_1, \dots, f^0_{\min A_1 - 1}$. $\min A_1$ is the *length of the beginning*. We say that a coloring of $\binom{A}{p}$ is *good* or that $\binom{A}{p}$ is *well-colored* if any two p-parameter sets with the same beginning have the same color.

Our proof proceeds by induction on p. The theorem for $p = 0$ reduces to the Hales–Jewett theorem. In the induction step let us assume the validity of the theorem for $p - 1$ and arbitrary A, n, t. Ramsey's theorem for parameter sets then follows from the following two claims.

Claim A (Sufficiency of well-coloring). *The following two statements are equivalent*:

(1) *For every n, t, there exists N such that if $\boldsymbol{C}$ is an N-parameter set and $\binom{C}{p}$ is colored by t colors then there exists $\boldsymbol{B}' \in \binom{C}{n}$ which is well-colored; we denote this by $N \xrightarrow[\text{good}]{} (n)^p_t$.*

(2) $N \rightarrow (n)^p_t$ (*for parameter sets*).

Claim B (Good well-coloring lemma). *For any n, m, t there exists an N such that if $\boldsymbol{C}$ is an N-parameter set then the following holds: For every t-coloring of $\binom{C}{p}$ there exists an n-parameter set $\boldsymbol{B} \in \binom{C}{n}$ determined by $(f^0, A_1, \dots, A_n)$ such that if $\boldsymbol{A}$ and $\boldsymbol{A}'$ are p-parameter subsets of $\boldsymbol{B}$ with the same beginning of length at most* $\min A_m$, *then $\boldsymbol{A}$ and $\boldsymbol{A}'$ have the same color.*

We denote this statement by $N \xrightarrow[\text{good},m]{} (n)^p_t$. Clearly, setting $n = m$ we obtain $N \xrightarrow[\text{good}]{} (n)^p_t$.

Proof of Claim B. We use the induction assumption for the Graham–Rothschild theorem for $p - 1$ (for arbitrary choice of A, n, m, t) and then (for given p, t, and n arbitrary) we proceed by induction on m. We allow $m = 0$ in which case the statement is trivial ($N = n$). Set $|A| = a$.

In the induction step, let t, n be given.

$$N_1 = \mathrm{GR}(a + 1, p - 1, t^{a^m}, n - m)\,,$$
$$N_2 = N_1 + m + 1\,,$$
$$N \xrightarrow[\text{good},m]{} (N_2)^p_t\,.$$

We prove

$$N \xrightarrow[\text{good},m+1]{} (n)^p_t\,.$$

Without loss of generality, set $\boldsymbol{C} = A^N$ and let c be a fixed t-coloring of $\binom{C}{p}$.

First apply the definition of N and let $\boldsymbol{B}'$ be an N_2-parameter subset of $\boldsymbol{C}$ determined by $f^0, A_1, \dots, A_{N_2}$ such that any two p-parameter sets in B' with the same beginning of length at most $\min A_m$ have the same color (cf. the definition of $\xrightarrow[\text{good},m]{}$). Set $\lambda = \min A_{m+1}$.

Now consider the set $\mathcal{A}$ of all p-parameter subsets $(g^0, H_1, \dots, H_p)$ of $\boldsymbol{B}'$ for which $\min H_1 = \lambda$. Every such p-parameter set $\boldsymbol{A}$ (over A) may be considered as a $(p - 1)$-parameter set $\bar{A}$ determined by $(\bar{g}^0, H_2, \dots, H_p)$ over (the enriched alphabet) $A \cup \{a+1\} = \bar A$ by simply setting:

$$\bar g_i^0 = \begin{cases} g_i^0 & \text{if } g_i^0 \text{ is defined}, \\ a+1 & \text{if } i \in \Pi_1 - \{\lambda\}. \end{cases}$$

Now define a coloring of $(\bar A^{N_1}_{p-1})$ by

$$\bar c(A') = (c(A); A \in \mathcal{A}, \bar A = A').$$

Given $A' \in (\bar A^{N_1}_{p-1})$ there are (exactly) a^m parameter sets A for which $\bar A = A'$ and thus we apply the definition of N_1 to obtain a $(n-m)$-parameter set $\bar B$ in $(\bar A^{N_1}_{p-1})$ which is $\bar c$-monochromatic. Let $\bar B$ be determined by $(\bar f^0, \bar A_{m+1}, \ldots, \bar A_n)$. Define the n-parameter set B given by $(h^0, \sum_1, \ldots, \sum_n)$ by

$$\begin{aligned} &h_i^0 = \bar f_i^0 \quad \text{if } \bar f_i^0 \text{ is defined}, \\ &\sum_i = A_i \quad \text{for } i < m, \\ &\sum_m = \{i; \bar f_i^0 = a+1\} \cup \{\lambda\}, \\ &\sum_i = \bar A_i \quad \text{for } i = m+1, \ldots, n. \end{aligned}$$

It can be easily checked (mainly from the definition of $\bar c$) that any two p-parameter subsets of B with length of beginning $\leq \min A_{m+1}$ have the same color (in the coloring c). □

Proof of Claim A. Clearly (2)$\Rightarrow$(1). The reverse implication follows clearly from the following truncated version of the Hales–Jewett theorem (isolated first by Voigt 1980) which thus plays the role of the pigeonhole principle in Proof II of Ramsey's theorem. This we state as Claim C.

First let us introduce the following: Let $A^{\leq n}$ denote the set $\bigcup_{m=1}^{n} A^m$. If B is an n-parameter set determined by $(f^0, A_1, \ldots, A_n)$ then a *partial point of* B is a point x of A^m for an $m \leq n$ which would become a point of B if we define its coordinates for $i > m$ by suitable constants (either f_i^0 or, say, 1). Now we can formulate the following.

Claim C (Hales–Jewett's pigeonhole). *Let A, t, n be fixed. Then there exists $N = \mathrm{HJ}^*(n,t)$ such that for every t-coloring of $A^{\leq N}$ there exist an n-parameter subset B of A^N such that the set of all partial points in B is monochromatic.*

Clearly, in order to finish the proof of Theorem 4.3 it suffices to prove Claim C.

Proof of Claim C. Without loss of generality assume $t = 2$ and denote by $N = N(n_1, n_2)$ a positive integer such that for every 2-coloring of $A^{\leq N}$, there exists an $i \in \{1, 2\}$ and an n_i-parameter subset B of A^N, whose partial points are monochromatic. We proceed by induction on $n_1 + n_2$. In the induction step we prove

$$N(n_1, n_2) \leq \mathrm{HJ}(A, 2^x, 1) + N = M + N\ ,$$

where we put $x = 1 + |A^{<N}|$ and $N = \max\{N(n_1 - 1, n_2),\ \mathrm{HJ}^*(n_1, n_2 - 1)\}$. Towards this end, let c be a 2-coloring of $A^{<M+N}$. Define a (2^x)-coloring c' of A^M by

$$c'(f) = (c(f, g)\colon g \in A^{\leq N}) \cup (c(f))\ .$$

Let $\boldsymbol{B}'$ be a c'-monochromatic line in A^M. In particular, $\boldsymbol{B}'$ is a c-monochromatic line, e.g., let $c(f) = 1$ for any $f \in \boldsymbol{B}'$. Now define the coloring c'' by

$$c''(g) = c(f, g)$$

for some (any) $f \in \boldsymbol{B}'$.

By assumption there exists $\boldsymbol{B}'' \in \binom{A^N}{m}$ for which $\boldsymbol{B}''^{\leq m}$ is c'' monochromatic. If $m = n_2$ and $\boldsymbol{B}''^{\leq m}$ is colored by 2, we are done since all partial points of $\boldsymbol{B}''$ have to be colored by 2.

If $m = n_1 - 1$ and $\boldsymbol{B}''^{\leq m}$ is colored by 1 then we combine B'' to a $\boldsymbol{B} \in \binom{A^{N+M}}{n_1}$ with $\boldsymbol{B}^{\leq n_1}$ colored by 1. □

Let us remark at the end of this section that we could define n-parameter sets in a slightly more technical way: Let A be a non-empty set and B (a possibly empty) subset of A.

An *n-parameter set* (with respect to B) in A^N is a non-empty set $\boldsymbol{A}$ of the following form: there are non-empty disjoint subsets $A_1, \ldots, A_n$ of $\{1, 2, \ldots, N\}$ and a point $f^0 = (f_1^0, \ldots, f_N^0)$ of B^N such that $\boldsymbol{A}$ is just the set of all points $f \in A^N$ which satisfy

$$f_i = \begin{cases} f_i^0 & \text{for } i \not\in \bigcup_{j=1}^{n} A_j\ , \\ f_{i'} & \text{for } i, i' \in A_j,\ j = 1, \ldots, n\ . \end{cases}$$

(Thus we just assume that the "affine shift" belongs to B^N.) Note that if $B = \emptyset$ then this definition implies that $(A_1, \ldots, A_n)$ forms a partition of $\{1, \ldots, N\}$.

For fixed B, A, denote again by $\binom{A}{p}$ the set of all p-parameter sets (with respect to B) in $\boldsymbol{A}$.

Note that the above proof actually gives the following slightly stronger result (also proved by Graham and Rothschild 1971).

Theorem 4.4. *Let n, t, p be positive integers, A a finite alphabet, B a subset of A. Then there exists $N = \mathrm{GR}(A, B, p, t, n)$ with the following property: If $\boldsymbol{C}$ is an N-parameter set (w.r.t. B) and $\binom{C}{p}$ is colored by t colors then there exists an n-parameter set $\boldsymbol{B}'$ for which $\binom{B'}{p}$ is a monochromatic set.*

We shall make use of this refinement in section 4.4.

In fact, Graham and Rothschild actually prove a version of the parameter set theorem in which a finite permutation group is allowed to act on the entries of the elements (see Graham and Rothschild 1971 for details).

4.3. *Vector and affine spaces*

The present author thinks that the Graham–Rothschild theorem 4.3 on parameter sets was the first theorem of the "new Ramsey theory" age – the first theorem with a proof which displayed the richness of the field of structural extensions of Ramsey theory and attracted so much attention to it. Moreover, it led *directly*, using a convenient formalism, to a solution of the following problem which has been responsible for much of the development in the sixties:

Rota's conjecture. The analogue of Ramsey's theorem holds for finite vector spaces.

This was a nice and natural problem since there are many similarities between the structure of subsets of a set and the structure of subspaces of a finite vector space. To be more specific, let $F = \mathrm{GF}(q)$ be a finite field with q elements, V an n-dimensional vector space over F. Denote by $\binom{V}{p}$ the set of all p-dimensional vector subspaces of V.

Using this, Rota's conjecture looks formally similar to Ramsey's theorem: For every finite field F and every choice of positive integers p, t, n there exists $N = N(p, t, n)$, such that for every t-coloring of $\binom{V}{p}$ for an N-dimensional vector space V over F, there exists an n-dimensional subspace U such that $\binom{U}{p}$ is monochromatic.

Let us remark that although points of V (i.e., $\binom{V}{0}$) form just the set F^n (or $\binom{F^n}{0}$), i.e., the set of all 0-parameter sets) this analogy does not hold for $p > 0$: Every p-parameter set may be regarded as a p-dimensional vector subspace of V; however, this does not hold in the other direction. (Figure 4.2 schematically illustrates a 2-dimensional subspace of V; compare it with fig. 4.1.) Motivated by this analogy 1-parameter sets are called *combinatorial lines*, etc.

Note that the size of $\binom{V}{p}$, denoted by $\left[{n \atop p}\right]_q$, (a *Gaussian coefficient*) shares many

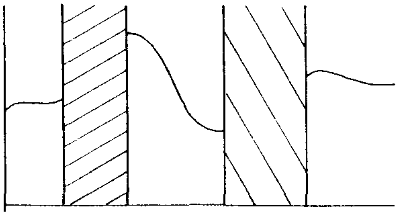


Figure 4.2.

properties with binomial coefficients. In particular, a formally similar definition is:

$$\left[{n \atop p}\right]_q = \frac{(q^n - 1)(q^n - q)\cdots(q^n - q^{n-p+1})}{(q^p - 1)(q^p - q)\cdots(q^p - q^{p-1})} .$$

However, these analogies were not of much help in solving the problem and it took nearly 10 years before Rota's problem was solved.

Theorem 4.5 (Graham et al. 1972). *For every finite field F and any choice of nonnegative integers p, n, t there exists $N = \mathrm{GLR}(p, n, t)$ such that for every t-coloring of $\binom{V}{p}$ for an N-dimensional vector space V over F, there exists an n-dimensional subspace U such that the set $\binom{U}{p}$ is monochromatic.*

The original proof relied on categorical formalism and was not easy. A simpler proof was given by Spencer 1979) while the proof along the lines given in the preceding section (via good colorings) was given by Deuber and Voigt (1982).

A companion result deals with affine spaces and is also due to Graham et al. (1972).

The mutual interplay of parameter sets and of vector and affine spaces, or to put it in other words, of combinatorial and "geometrical" lines, is very interesting. While for the coloring results (Ramsey type theorems) the algebraical structure involved in Rota's conjecture presented difficulty and thus this conjecture has been solved later than its combinatorial counterpart (i.e., parameter sets) for density theorems this has been the other way around. The density version of Theorem 4.5 was established by the following:

Theorem 4.6 (Fürstenberg and Katznelson 1985). *For every finite field F, positive integer n and a positive real $\varepsilon > 0$ there exists a positive integer $N = \mathrm{KF}(\varepsilon, F, n)$ with the following property: If X is a set in an N-dimensional vector space of size at least $\varepsilon \cdot |F|^N$ then X contains an n-dimensional vector space.*

But the density of Hales–Jewett's theorem appeared much harder. Although the boolean case was settled (Brown and Buhler 1982, Rödl 1982) and also some reductions were known (Brown and Buhler 1984) the lack of symmetries of cubes A^N left the problem open for several years (well after the first version of this paper was written). The problem was solved in the beginning of 1990:

Theorem 4.7 (Density of Hales–Jewett's theorem; Fürstenberg and Katznelson 1991). *For every alphabet A, positive integer n and positive real $\varepsilon > 0$ there exists a positive integer $N = N(\varepsilon, A, n)$ with the following property: if X is a subset of A^N of size $\varepsilon \cdot |A|^N$, then X contains an n-parameter set.*

The difficulty of handling combinatorial lines in density questions is a bit surprising since in most Ramsey-type questions both combinatorial- and vector-

(affine)-structures are usually closely related. In fact there exist just a few instances of problems where these two structures differ.

One of them is so-called *selectivity*, see, e.g., Erdős et al. (1984b). Namely, if the points of a large dimensional vector-space are colored by an arbitrary number of colors then one of the n-dimensional subspaces is either monochromatic or totally multicolored (meaning that no two points have the same color).

A similar selective property holds for Van der Waerden's theorem (as shown by Erdős and Graham 1980) and for Rado's theorem (as shown by Lefmann 1986).

However, this does not hold for Hales–Jewett cubes: Consider an arbitrary Hales–Jewett cube A^N, $A=\{1,2,\dots,a\}$, $a\geq 3$. Put $c(1)=1$ and $c(2)=c(3)=\cdots=c(a)$. Define a coloring c of A^N by

$$c((x_1,\dots,x_N))=\sum_{i-1}^{N} c(x_i)\cdot 3^i\ .$$

One sees easily that while there is no monochromatic line, in every line two points get the same color. Also this instance indicates the lack of symmetries of Hales–Jewett cubes.

The full discussion of coloring patterns which may occur in unrestricted colorings is subject to canonical Ramsey theory, named after the Canonical partition lemma due to Erdős and Rado (1952).

Theorem 4.8 (Canonical Lemma, Erdős and Rado 1952). *For every choice of positive integers p, n there exists $N(p,n)$ such that for every set X, $|X|\geq N(p,n)$ and for every coloring $c:\binom{X}{p}\to\mathbb{N}$ (i.e. coloring by arbitrary many colors) there exists a set $Y\subseteq X$, $|Y|\geq n$, such that the coloring c restricted to the set $\binom{Y}{p}$ is canonical. Here a coloring c is said to be canonical if there exists a set $\omega\subseteq[p]$ such that*

$$c(x_1<\cdots<x_p)=c(x'_1<\cdots<x'_p)\ \textit{iff}\ x_i=x'_i\quad \textit{for}\ i\in\omega\ .$$

(Thus there are 2^p canonical coloring patterns.)

Coloring patterns for Hales–Jewett cubes were determined by Prömel and Voigt (1983), see also the survey Prömel and Voigt (1985).

Erdős and Rado's proof yields the upper bound of order t_{2p} (while obviously $N(p,n)\geq r(p,n-1,n)$). Recently Lefmann and Rödl attacked the problem of estimating the numbers $N(p,n)$. They solved the problem for pairs (Lefmann and Rödl 1993, 1995) and very recently Shelah (1994) solved the problem for every p by showing that both upper and lower bound use $p-1$ exponentiations. For an elegant proof of the Erdős–Rado Canonization lemma see also Rado (1986).

Due to space limitations we cannot describe these interesting results in greater detail. This chapter belongs to the "Aspects" part of this Handbook of Combinatorics but we are unable to cover even one aspect of a single theory.

4.4. Finite union theorem and other variations

Several particular cases of partition regular systems of equations were studied in their own sake. Apart from Van der Waerden's theorem this is particularly true of the Finite union theorem.

Theorem 4.9 (Rado 1933, Folkman, Sanders 1968). *For every choice of positive integers n, t, there exists $N = \mathrm{FU}(n, t)$ with the following property: For every partition of $\{1, \ldots, N\}$ into t classes one of the classes contains n distinct numbers together with all their sums.*

The theorem gets its name after an equivalent formulation in terms of partitions of subsets.

Theorem 4.10 (Finite union theorem). *For every choice of positive integers n, t, there exists N with the following property: If X is a set of size at least N and if the power set 2^X is partitioned into t classes then in one of the classes we can find n disjoint sets, together with all their (non-empty) unions.*

Let us remark that a weaker version of this result is perhaps chronologically the earliest instance of a valid Ramsey-type statement (not counting Dirichlet's pigeonhole principle): Hilbert (1892) published the following result (which we formulate here in modern terms).

Theorem 4.11. *For any positive integers n, t, there exists $N = H(n, t)$ such that for every partition of the power set lattice 2^X for $|X| \geq N$ into t classes, one of the classes contains a sublattice isomorphic to 2^n.*

Proof. Apply the Hales–Jewett theorem for n and t to show that $H(n, t) \leq \mathrm{HJ}(n, t)$. □

(In fact a direct approach here yields a much better result, see Brown et al. 1985 and references given there.)

Let us remark that one can apply the Hales–Jewett theorem to prove a similar theorem not only for lattices in general but for all varieties of lattices, e.g., modular lattices, see Ježek and Nešetřil (1983) and Prömel and Voigt (1981) for details.

The Finite union theorem may also be proved from the Graham–Rothschild theorem for parameter sets. Let us give a simple proof as an application of Van der Waerden's theorem only.

Proof of Theorem 4.9. For simplicity we prove the statement for $t = 2$ in the following form: For any pair n_1, n_2 of positive integers there exists $\mathrm{FU}(n_1, n_2) = N$ with the following property: For every partition $\{1, \ldots, N\} = a_1 \cup a_2$, either a_1

contains an n_1-set together with all sums or a_2 contains an n_2-set together with all sums.

We proceed by induction on $n_1 + n_2$, and in the inductive step we show that

$$\mathrm{FU}(n_1, n_2) \leq \mathrm{VW}(1 + \mathrm{FU}(n_1, n_2 - 1), 2) .$$

To this end, set $\mathrm{FU}(n_1, n_2 - 1) = \mathrm{FU}(n_1 - 1, n_2) = M$, $\mathrm{VW}(1 + M, 2) = N$ and let $\{1, \ldots, N\} = a_1 \cup a_2$ be a fixed partition. Then there exists a_0 and $d > 0$ with all a_0, $a_0 + d, \ldots, a_0 + Md$ in one of the classes, say a_2.

Now consider the set $\{d, 2d, \ldots, Md\}$ under the given partition. Then either there exists an n_1-set together with all sums in a_1 and we are done, or there exists an $(n_2 - 1)$ set $dx_1, \ldots, dx_{n-1}$ such that all its sums are in a_2. But then d, $dx_1, \ldots, dx_{n-1}$ is a desired n_2-set. □

The Finite union theorem has always been related to (a version of) the Hales–Jewett theorem and thus the existence of a good (even of a primitive recursive) bound presented a problem. Using Shelah's proof one now has such a bound (of the order of the sixth-function of the hierarchy). However, using direct methods one can get bounds of order f_3 (the tower function), see Taylor (1981), Nešetřil and Rödl (1983b) and Graham and Rödl (1987).

The Finite union theorem has a countable analogue, namely Hindman's (1974) theorem. This is one of the few known partition regular infinite systems of equations, and in fact for infinite systems of equations one cannot hope for a statement similar to Rado's theorem as shown recently by Deuber et al. (1995). Hindman's theorem and its ultrafilter proof due to Glazer led to an important line of research. Summarizing, there are at least three possible approaches to these coloring problems: using topological dynamics (e.g., the linear Van der Waerden theorem proved in Fürstenberg and Weiss 1978), using Stone–Čech compactification of the corresponding structures (Bergelson and Hindman 1988), and using theory of ultrafilters (see, e.g., Carlson 1988). Although these approaches are related several of these results were obtained independently and in different context. And there were found several "master theorems". One of them is certainly the following very recent result which was conjectured by Fürstenberg and proved by Bergelson and Leibman (1995).

Theorem 4.12 (Polynomial Van der Waerden Theorem). *Let $p_1, \ldots, p_k$ be polynomials with rational coefficients taking integer values on integers and satisfying $p_i(0) = 0$ for $i = 1, \ldots, k$. Then for every finite partition of $\mathbb{N}$ and for every choice of numbers $v_1, \ldots, v_k$ there exist integers n, u and d such that $u + p_i(d)v_i$, for $i = 1, \ldots, k$ belong to the same class of the partition.*

(Choose $p_1(n) = \cdots = p_k(n) = n$, $v_i = i$ to get Van der Waerden's theorem.)

This theorem is proved by a transfinite induction over countable ordinals thus sharing some similarities with section 3.3. Perhaps there is a deeper connection here yet to be discovered.

Graham–Rothschild's theorem has many applications. For example if $A = \{1, 2\}$ then we obtain the Ramsey property of finite Boolean algebras since the n-parameter word corresponds to a Boolean subalgebra isomorphic to 2^n. In the terminology of section 5 this proves that the finite Boolean algebras form a Ramsey class.

We get another interesting case if we consider $B = \emptyset$, $A = \{1, 2\}$ (in Theorem 4.4.). As remarked earlier, in this case the moving coordinates of an n-parameter word form a partition into n classes and it is not difficult to check that A is a p-parameter subset of an n-parameter set $\boldsymbol{B}$ iff the partition which corresponds to $\boldsymbol{B}$ is a refinement of the partition which corresponds to A. Thus in this case we obtain the following result which is of independent interest: Let $\mathrm{Eq}(X)$ denote the set of all equivalences on X. We set $E \leqslant E'$ if the equivalence E' (as a relation) contains E (i.e., if E is a refinement of E'). Given an equivalence E and a positive integer p, denote by $\binom{E}{p}$ the set of all equivalences with p classes which are coarser than E. Then we have the following.

Theorem 4.13 (Dual Ramsey theorem). *Let p, n, t be positive integers. Then there exists $N = \mathrm{DR}(p, t, n)$ with the following properties: If X is a set with at least N vertices and c is a t-coloring of $\binom{\mathrm{Eq}(X)}{p}$ then there exists an equivalence E with n classes such that $\binom{E}{p}$ is monochromatic.*

Proof. Apply the Graham–Rothschild theorem 4.4 for $B = \emptyset$, $A = \{1, 2\}$ for p and n as above. □

Although this is given and proved in Graham and Rothschild (1971) and although one leitmotiv of Leeb's (1973) work was his interpretation of parameter sets as the dual category to the category of all finite sets, the tempting idea of a dual Ramsey theorem has been rediscovered several times (see Nešetřil and Rödl 1980 and Carlson and Simpson 1984). Carlson and Simpson (1984) proved an important infinite topological dual Ramsey theorem which may be viewed as a culminating result for parameter-set-type theorems; see Carlson and Simpson (1990) for a survey of this development.

There are other applications of the Graham–Rothschild theorem. For example, we already obtain from the Hales–Jewett theorem the following (known as the Gallai–Witt theorem).

Theorem 4.14. *For every t, n and m and for every t-coloring of the m-dimensional integer lattice points, there exists a monochromatic homothetic copy of $\{1, \ldots, n\}^m$. Explicitly there exists $(a_1, a_2, \ldots, a_m)$ and $d > 0$ such that all points of the form*

$$(a_1 + i_1 d, a_2 + i_2 d, \ldots, a_m + i_m d), \quad i_1, \ldots, i_m \in \{0, \ldots, n-1\}$$

are monochromatic.

Proof. Let t, m, n be given. Set $\bar{A} = \{0, 1, \ldots, n-1\}$, $A = \bar{A}^m$. Thus $A \subseteq \mathbb{Z}^m$. Using the Hales–Jewett theorem set $N = \mathrm{HJ}(A, t, n^m)$. Choose positive integers $k_1, \ldots, k_N$ so that the mapping

$$f : A^N \to \mathbb{Z}^m$$

defined by

$$f(x_1, \ldots, x_N) = \sum_{i=1}^{N} k_i x_i$$

is injective (this can be easily guaranteed; e.g., set $k_i = (N+1)^i$). Now any t-coloring of $\mathbb{Z}^m$ induces a t-coloring of A_t^N which contains a monochromatic line L determined by $g_0 \in A^N$ and $I \subseteq \{1, \ldots, N\}$. The image $f(L)$ of L induces a configuration in $\mathbb{Z}^m$ which is *homothetic* to $\{0, 1, \ldots, n-1\}^m$. Explicitly, this configuration is determined by $a_1, a_2, \ldots, a_m$ and d where

$$(a_1, \ldots, a_m) = \sum_{i=1}^{N} k_i g_0(i) \quad \text{and} \quad d = \textstyle\sum_{i \in I} k_i . \qquad \square$$

It follows from this that for every finite subconfigurations K of $\mathbb{Z}^m$ and for every finite coloring of $\mathbb{Z}^m$ one can find a monochromatic configuration which is homothetic to K.

However, the situation drastically changes if we insist that we find a monochromatic configuration which is congruent to a given one. This was the starting point of *Euclidean Ramsey theory* which deals with the following problem: Which finite configuration (up to a Euclidean motion) can we always find monochromatically in any finite coloring of the points in $\mathbb{E}^n$, where n is sufficiently large depending on the number of colors?) (Such configurations are called *Ramsey*.)

This problem is far from being solved and several partial results were obtained in a series of papers by Erdős et al., see, e.g., Erdős et al. (1975a). Recently the progress has been quick: Frankl and Rödl (1986) proved that every triangle and, more generally, that every nondegenerated simplex (Frankl and Rödl 1990) is Ramsey. This has been quickly generalized to trapezoids, pentagons and other notoriously difficult small configurations by Kříž (1991, 1992a). Kříž also applied deep methods of algebraic topology to this problem Kříž (1991, 1992b). Other recent striking results are contained in Bourgain (1986) and Matoušek and Rödl (1995) and the description of this recent development is given in Graham (1990, 1995). See also chapter 17 in this Handbook.

Another problem related to the multidimensional generalization of Van der Waerden's theorem is the density problem.

Theorem 4.15 (Fürstenberg and Katznelson 1978). *Let $\varepsilon > 0$, let d be a positive integer and X a subset of $\mathbb{Z}^d$ with a positive upper density. Then X contains a homothetic copy of $\{1, \ldots, n\}^d$ for every positive integer n.*

The proof of Fürstenberg and Katznelson used infinitary methods of ergodic theory. No combinatorial proof is known (even for this "simple" result).

Other geometric density results motivated by Van der Waerden's theorem were proved by Pomerance (1980) and Bourgain (1986).

A final comment on density problems: One may ask whether a density theorem is valid for all other partition regular systems of equations $Ax = 0$.

The answer is negative as has been observed by Frankl et al. (1988).

Theorem 4.16. *For a partition regular system $Ax = 0$ which has a solution consisting of pairwise distinct entries, the following two statements are equivalent:*

(1) $A \cdot \mathbf{1} = 0$ *where* $\mathbf{1} = (1, \ldots, 1)^{\mathrm{T}}$.

(2) *For every set X of positive integers with a positive upper density the system*

$$Ax = 0$$

has a solution in X which consists of pairwise distinct entries.

Proof. This follows from Szemerédi's theorem:

(1)$\Rightarrow$(2) Let $x = (x_1, \ldots, x_n)$ be a solution of $Ax = 0$ with all x_i distinct. Put $N = \max x_i$ and let $c + jd$, $j = 0, \ldots, N$ be an arithmetic progression in X (by virtue of Szemerédi's theorem). Now if $A\mathbf{1} = 0$ then also (by linearity)

$$Ay = 0\,,$$

where $y = c \cdot \mathbf{1} + d \cdot x$. All entries of y have the form $c + x_i d$, are distinct, and belong to X.

(2)$\Rightarrow$(1) Set $N > \sum_{i,j} |a_{ij}|$ (where a_{ij} is a general term of matrix A). Consider the set $X = \{Ny + 1\colon y = 1, 2, \ldots\}$. X has upper density $1/N$. Applying (2), let $x = (x_1, \ldots, x_n)$ be a solution of $Ax = 0$ in X. Thus each $x_j = N \cdot y_j + 1$ and

$$0 = \sum_{j=1}^{n} a_{ij} x_j = \sum_{j=1}^{n} a_{ij}(Ny_j + 1) = N \sum_{j=1}^{n} a_{ij} y_j + \sum_{j=1}^{n} a_{ij}$$

for every $i = 1, \ldots, m$.

This implies $\sum_{j=1}^{n} a_{ij} = 0$ for every i which in turn is just $A\mathbf{1} = 0$. □

5. Ramsey classes

5.1. Induced theorems

The examples of full Ramsey theorems which we gave in the previous section share a common "Ramsey pattern". As we remarked earlier it is difficult to enrich this list by essentially new examples. It seems that there somehow is a limited supply of finite structures (from "real life") with the full Ramsey property. Nevertheless, there is a rich theory of structural extensions of Ramsey theory which is presently perhaps one of the most active areas of the subject.

By now we all tend to consider this as a natural state of the art but the situation in the early 70s was far less transparent and the new concepts emerged slowly. These concepts then influenced the further development in a profound way.

It seems that one of the significant turns appeared in the late 60s when Erdős, Hajnal and Galvin started to ask questions such as which graphs arrow, say, a triangle. Perhaps the essential parts of the whole development can be illustrated with this particular example.

We say that a graph $G = (V, E)$ is *t-Ramsey* for the triangle (i.e., K_3) if for every t-coloring of E, one of the classes contains a triangle. Symbolically we denote this by $G \to (K_3)^2_t$. Ramsey's theorem gives us $K_6 \to (K_3)^2_2$ (and $K_{r(2,t,3)} \to (K_3)^2_t$).

But there are other essentially different examples. For example, a 2-Ramsey graph for K_3 need not contain K_6.

This was shown by Graham (1968) who constructed an optimal graph with this property: The graph $K_3 + C_5$ (depicted in fig. 5.1) is the smallest graph G for which $G_5 \to (K_3)^2_2$ and which does not contain K_6.

Yet $K_3 + C_5$ contains a K_5 and subsequently, Van Lint, Graham and Spencer constructed a graph G_4 not containing K_5, with $G_4 \to (K_3)^2_2$. Until recently, the smallest such an example was due to Irving (1973) and had 18 vertices. (It arises from a graph on 17 vertices which proves $17 \not\to (4)^2_2$, by adding one vertex of degree 17.) Very recently two more constructions appeared by Erickson (1993) and Bukor (1994) who found examples with 17 and even 16 vertices (both of them using properties of Graham's graph).

The next question which was asked is whether there exists a K_4-free graph G_3 for which $G_3 \to (K_3)^2_2$. This question proved to be considerably harder and it is possible to say that it has not yet been solved satisfactorily.

The existence of a K_4-free graph G which is 2-Ramsey for K_3 was settled by Folkman (1970). Folkman's proof is complicated and the constructed graph is astronomically large. The same is true for simpler constructions of these graphs as provided by Nešetřil and Rödl (1975a, 1981).

Perhaps just to be explicit Erdős (1975) asked whether there exists a graph G_3 with $< 10^{10}$ vertices. This question proved to be very accurate and, building on earlier partial results of Szemerédi, Frankl and Rödl, it was shown by Spencer (1988) that there exists such a graph G_3 with $300 \cdot 10^9$ vertices.

The proof of this statement is probabilistic. But the probabilistic methods were

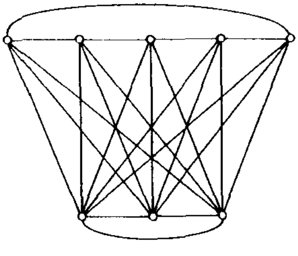


Figure 5.1.

not only applied to get various bounds for Ramsey numbers. Recently also the Ramsey properties of the random graph $G_{N,p}$ were analysed and the problem has been recently solved by Rödl and Ruciński (1993, 1995). More preciesely, Rödl and Ruciński determined (up to a multiplicative constant) for every graph G the threshold probability $p = p(G, N)$ for the property that almost every graph $G_{N,p}$ is Ramsey for G: $G_{N,p} \rightarrow (G)^2_2$. These results have further consequences which will be mentioned later.

Let us return to the beginning of the 70s. The situation was of course much less clear. On one side there were several deep and difficult results (Szemerédi's on $r_4(n)$; the Graham–Rothschild theorem, Folkman's theorem), on the other side the proofs were difficult and provided no insight. There were also strange difficulties: Rota's conjecture could be solved only for points, Folkman's proof did not work for 3-colorings, etc.

By means of the intensive research in the beginning of the 70s all these obstacles were resolved. An important part of this development was the fact that the theory got the proper footing in the work of Leeb (1973), Nešetřil and Rödl (1975a, 1978a), Deuber and Rothschild (1976) and others.

We shall give the main definitions.

Let K be a class of objects which will be denoted by A, B, $C, \ldots$ (such as graphs, subsets of integers, spaces).

Let K be endowed with isomorphisms and subobjects (such as subgraphs, subconfigurations, subspaces).

Given objects A and B, denote by $\binom{B}{A}$ the set of all subobjects of B which are isomorphic to A.

We say that object C is *(t, A)-Ramsey* for object B if for every t-coloring of the set $\binom{C}{A}$ there exists a subobject B' of C which is isomorphic to B such that the set $\binom{B'}{A}$ is monochromatic.

We denote this by $C \rightarrow (B)^A_t$. (It will be understood from the context in which class K we work.)

For $A \in K$ we say that the class K has the *A-Ramsey property* if for every object B of K and every positive integer t there exists C of K such that

$$C \rightarrow (B)^A_t .$$

In the extreme case where K has the A-Ramsey property for each of its objects A we say that K is a *Ramsey class.*

The *Ramsey problem for a class K* ("prototype theorem" in Nešetřil and Rödl 1975a) consists of describing all those objects A for which K for which K has the A-Ramsey property.

These definitions proved to be useful mainly because they provide a concise language for various extensions of Ramsey's theorem.

After reading the initial sections of this chapter one perhaps sees that these definitions are natural for most of the theorems which were introduced.

Every full Ramsey-type theorem induces a particular A-Ramsey property since it assures that *every* sufficiently large object is Ramsey. Thus we have:

(1) Set: The class of all finite sets together with subset inclusion forms a Ramsey class.

(This is just Ramsey's theorem.)

(2) Δ (simplicial class): The class of totally-ordered finite sets together with 1–1 monotone mappings as subobjects is a Ramsey class.

(This is again another formulation of Ramsey's theorem.)

(3) Compl: The class of all complete graphs is a Ramsey class.

(This is yet another reformulation of Ramsey's theorem.)

(4) Par(A): The class of all parameter sets over A is a Ramsey class (the Graham–Rothschild theorem 4.3.)

(5) AProg: The class of all finite subsets of integers and affine embeddings has the A-Ramsey property iff $|A| = 1$. (Here if $A = \{a_1 < a_2 < \cdots < a_m\}$ and $B = \{b_1 < b_2 < \cdots < b_n\}$ are two sets of integers then $f: A \rightarrow B$ is an affine embedding if there exists $d > 0$ such that $b_j = f(a_1) + j \cdot d$ iff $b_j = f(a_j)$, $1 \leq j \leq m$.)

This is the induced form of Van der Waerden's theorem, see Spencer (1988) and Nešetřil and Rödl (1976).

(6) Vect F: The class of all finite vector spaces over F is a Ramsey class.

(This is the Graham–Leeb–Rothschild theorem 4.10.)

One can formulate in this language further results which we obtained earlier. However, one does not have to look for full Ramsey theorems only. The weakening of the assumption to an *existence* of a Ramsey object proved to be both natural and useful: The Ramsey classes form a much richer variety than the (rather solitary) full Ramsey theorems.

Let us list some examples of Ramsey classes together with their (by now canonical) names.

Denote by Gra the class of all finite graphs together with induced subgraphs (as subobjects). Then we have the following solution of the Ramsey problem for graphs.

Theorem 5.1 (Deuber 1975a, Nešetřil and Rödl 1975b). Gra *has the A-Ramsey property* iff A *is either a complete graph or the complement of a complete graph* (*which will be called discrete*).

This extends results of Erdős et al. (1975b) and Rödl (1976) (the case $A = K_2$).

Proof. First we prove that Gra has the A-Ramsey property whenever A is complete or discrete. This is nontrivial and original proofs were difficult. However, one can derive this particular result from the Graham–Rothschild theorem on parameter sets (we follow the proof in Nešetřil and Rödl 1985).

Clearly it suffices to consider the case when A is complete. Set $|A| = p$.

We use the easy and well-known fact that every graph $G = (V, E)$ may be represented as a subgraph of the graph $(\mathcal{P}(X), \{[Y, Z]: Y \cap Z = \emptyset\})$ for sufficiently large X (it obviously suffices to assume $|X| \geq |E|$), say $|X| = n$. Moreover, this representation may be chosen such that a given order on V coincides with minimal elements of the sets which represent vertices.

However, it now suffices to apply the Graham–Rothschild theorem 4.4 for $B=\{1\}$, $A=\{1,2\}$ and $p=2$: Every n-parameter set in A^n gives a monochromatic subgraph (monotone) isomorphic to $(\mathcal{P}(X), \{[Y,Z]: Y\cap Z=\emptyset\})$.

The proof of the reverse implications uses the following technique developed in Nešetřil and Rödl (1975b).

Ordering lemma 5.2. *Let $G=(V,E)$ be a graph, $\leq$ an arbitrary linear ordering of the set V. Then there exists a graph $H=(W,F)$ with the following property: For every linear ordering $\preccurlyeq$ of W, there exists an embedding $f: G\rightarrow H$ which is a monotone mapping with respect to $\leq$ and $\preccurlyeq$. Explicitly: $[x,y]\in E$ iff $[f(x),f(y)]\in F$ and $x\leq y$ iff $f(x)\preccurlyeq f(y)$.*

Using the Ordering lemma we can prove the reverse implication in Theorem 5.1.

Let A be a graph which is noncomplete and nondiscrete. It follows that the vertex-set $V(A)$ of A may be ordered in two different ways, say $\leq_1$ and $\leq_2$ so that $(A,\leq_1)$ and $(A,\leq_2)$ are monotone non-isomorphic. Let $(A',\leq)$ be the disjoint union of $(A,\leq_1)$ and $(A,\leq_2)$, ordered in such a way that $\leq$ extends both $\leq_1$ and $\leq_2$.

Let $(B,\leq)$ be a graph guaranteed by the Ordering lemma for $(A',\leq)$. We claim that there is no C with $C\rightarrow(B)^A_2$: Let C be a graph. Fix an arbitrary linear ordering $\lhd$ of $V(C)$ and define a partition $\binom{C}{A}=a_1\cup a_2$ as follows:

$$\bar{A}\in a_1 \text{ iff } (\bar{A},\lhd|_{V(\bar{A})}) \text{ is monotone isomorphic to } (A,\leq_1),$$

$$\bar{A}\in a_2 \text{ otherwise.}$$

Using the ordering property of B we find that there is no homogeneous copy of B in C. □

Proof of the Ordering lemma. We use the edge-Ramsey property of Gra which we established in the first part of the proof.

Let G be any graph with a fixed linear ordering $\leq$ of its vertices. Let $\bar{G}$ be any graph with a linear ordering of its vertices such that both $(\bar{G},\leq)$ and $(\bar{G},\geq)$ contain an induced copy of G, monotone isomorphic to $(G,\leq)$. Put $\bar{G}=(V,E)$, $V=\{v_1<\cdots<v_n\}$. We may assume without loss of generality that $[v_i,v_{i+1}]\in E$ for $i=1,\ldots,n-1$ (for otherwise we may add to $\bar{G}$ a suitably ordered path from v_i to v_{i+1}).

Now according to the first part of the above proof of Theorem 5.1 there exists a graph $H=(W,F)$, $W=\{w_1<\cdots<w_N\}$, such that for every partition $F=F_1\cup F_2$ one of the classes contains an induced copy of $\bar{G}$ which is, moreover, monotone isomorphic to $\bar{G}$.

We claim that H has the ordering property for $(G,\leq)$. Towards this end let $\leq$ be a fixed linear ordering of W, $W=\{v_1<\cdots<v_n\}$. Let $\preccurlyeq$ be an arbitrary linear

ordering of W. Define a partition $a_1 \cup a_2$ of edges of H as follows:

$$[v_i, v_j] \in a_1 \text{ iff } v_i < v_j \text{ and } i < j ,$$

$$[v_i, v_j] \in a_2 \text{ otherwise} .$$

We get a monochromatic copy of $\bar{G}$ in one of the classes. From this it follows that for all edges $[v_i, v_j]$, either the orderings $<$ and $<$ coincide or they are inverse. By the definition of $\bar{G}$ this implies that in either case we get a monotone copy of G. □

Thus the ordering property may be deduced from the nonsingleton Ramsey property. However it is important that we are able to prove the ordering property independently of the Ramsey property for "almost" every structure, see Nešetřil and Rödl (1978b).

This is also mirrored by the fact that one can find small example graphs H with the ordering property (for G with n vertices one can choose H of order $n^2 \cdot \log n$; see Rödl and Winkler 1989; see also Brightwell and Kohayakawa 1993 and Nešetřil 1994 for recent related results).

For certain structures the validity of an induced Ramsey-type theorem was not easy to prove.

So it was for hypergraphs and (more generally) relational systems. This development culminated in the proof of Ramsey's Theorem for structures, proved independently by Abramson and Harrington (1978) and Nešetřil and Rödl (1977b). We shall state this theorem after introducing the necessary notions:

A *type* is a sequence $(n_\delta; \delta \in \Delta)$ of positive integers. A type will be fixed. A *structure* (set system) *of type* Δ is a pair $(X, \mathcal{M})$ where:

(1) X is a linearly ordered set (this ordering we call *standard*);

(2) $\mathcal{M} = (\mathcal{M}_\delta; \delta \in \Delta)$ and $\mathcal{M}_\delta \subseteq \binom{X}{n_\delta}$ for each $\delta \in \Delta$.

Given two structures $(X, \mathcal{M})$ and $(Y, \mathcal{N})$, $\mathcal{N} = (\mathcal{N}_\delta; \delta \in \Delta)$, a mapping $f: X \to Y$ is said to be an *embedding* if

(1′) f is 1–1 and monotone with respect to standard orderings;

(2′) For every $\delta \in \Delta$ and each subset $M \subseteq X$ we have

$$M \in \mathcal{M}_\delta \text{ iff } f(M) \in \mathcal{N}_\delta .$$

An *isomorphism* is an invertible embedding.

$(X, \mathcal{M})$ is a *substructure* of $(Y, \mathcal{N})$ iff the inclusion $X \subseteq Y$ is an embedding. Given two structures A and B denote by $\binom{B}{A}$ the set of all substructures of B which are isomorphic to A.

Denote by Soc(Δ) the class of all structures of type Δ together with substructures. Then we have the following result obtained by Nešetřil and Rödl (1977b, 1983a) and, independently, by Abramson and Harrington (1978).

Theorem 5.3 (Structural Ramsey theorem). Soc(Δ) *is a Ramsey class.*

Another way to view a structure $(X, \mathcal{M})$, $\mathcal{M} = (\mathcal{M}_\delta; \delta \in \Delta)$, is to consider a *structural coloring* s_X of the power set $\mathcal{P}(X)$ (i.e., a mapping $s_X: \mathcal{P}(X) \rightarrow S$) defined as follows: the color of a set M is the set of all δ for which $M \in \mathcal{M}_\delta$.

Embeddings are then just monotone mappings which preserve structural coloring. Explicitly, if s_X is a structural coloring of an ordered set X and s_Y is a structural coloring of an ordered set Y then $f: X \rightarrow Y$ is an *embedding* of (X, s_X) into (Y, s_Y) if f is monotone and 1–1 and for every set $M \subseteq X$, (2′) holds. The following diagram shows the embedding:

$$X \xrightarrow{f} Y$$

$$\begin{tikzcd}
\mathcal{P}(X) \arrow[rr] \arrow[dr, "s_X"'] & & \mathcal{P}(Y) \arrow[dl, "s_Y"] \\
& S &
\end{tikzcd}$$

If two structures (specified by structural colorings) are isomorphic then we say that they have the same *pattern*.

This definition of a structure by means of the structural coloring has some advantages, one of them being that a structure and its "complement" induce (up to a permutation of labels) the same pattern.

With this we can formulate the above theorem as follows.

Theorem 5.4 (Structural Ramsey theorem). *For every set X endowed with a structural coloring $s_X: \mathcal{P}(X) \rightarrow S$ and for every positive integer t there exists a set Y endowed with a structural coloring $s_Y: \mathcal{P}(Y) \rightarrow S$ such that for every t-coloring of $\mathcal{P}(Y)$ there exists a set X' such that s_Y induces on X' the same pattern as the pattern of X, s_X, and moreover, the two subsets of Y' with the same pattern have the same color.*

Let us compare Theorem 5.4 with Ramsey's theorem 1.1. By iterating Ramsey's theorem we can easily get the following result: For every t and n there exists N such that for every coloring of the power set $\mathcal{P}([N])$ by t colors there exists a subset X of $[N]$ of size n such that the color of a subset of X depends only on its size. Such a homogeneous set may be visualized by fig. 5.2.

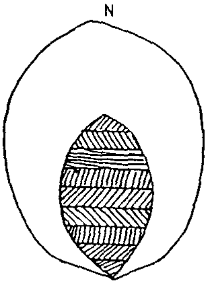


Figure 5.2.

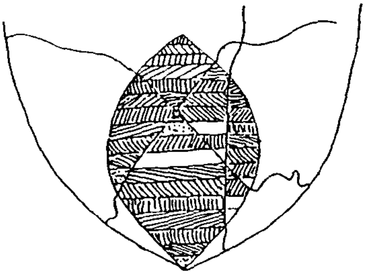


Figure 5.3

The structural Ramsey theorem 5.4 may then be visualized by the following more modern, somewhat Klee-type (Klee 1923) picture shown in fig. 5.3.

Theorems 5.1, 5.3 and 5.4 are called *induced Ramsey-type theorems*. For every structure with a valid Ramsey property we may consider an induced version (by enriching the objects by the addition of a structural coloring). All these problems have now been solved. For example, the induced theorem for Hales–Jewett has been established in Deuber et al. (1982) and the induced theorem for parameter sets, vector and affine spaces has been established by Prömel (1985).

The original proofs of these theorems were not easy. On the other hand as we shall see in the next section, the induced theorems present only the first step on the ladder of difficulty of Ramsey-type theorems. The general methods imply them all . . . .

5.2. Restricted theorems

We motivate this chapter by (the chronologically first) example of a restricted Ramsey problem (i.e., K_4-free Ramsey graphs for the triangle).

This problem was partially solved by Folkman (1970) and the full solution was obtained by Nešetřil and Rödl (1975a).

Theorem 5.5. *Let G be a graph not containing a complete graph K_k. Let t be a positive integer. Then there exists a graph H not containing a complete graph K_k such that for every t-coloring of edges of H we get a subgraph isomorphic to G with all its edges in one of the classes of the partition.*

Results of this type are called *restricted* Ramsey-type theorems. One should note that this result implies the induced Ramsey theorem for graphs. This is a general comment: *Restricted theorems imply induced theorems.*

We shall justify this claim by proving it for Theorem 5.5, for simplicity, say for $k = 3$: Thus, let G be a graph without triangles. Find a graph I without triangles, with two vertices a, b of I such that $[a, b]$ does not form an edge in I, and if we add the edge $[a, b]$ to I then we get a triangle.

It is easy to find such an I, e.g., a path of length 2 will do.

Now build a graph G' by amalgamating a copy of I to every pair x, y of independent points of G in such a way that x is identified with a and y with b.

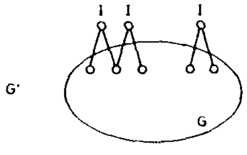


Figure 5.4.

See the schematic in fig. 5.4. Obviously G' does not contain a triangle.

It is easy to see that if H is a triangle-free graph which for every t-coloring of its edges contains a subgraph isomorphic to G' then also

$$H \rightarrow (G)_t^{K_2}$$

in the class of all graphs and (induced) subgraphs.

This leads to the following definition: Let Forb(K_k) be the class of all finite graphs which do not contain K_k with induced subgraphs as subobjects. (Thus Forb(K_k) is a subclass of Gra.)

For classes Forb(K_k), the solution of the Ramsey problem is provided by the following result of Nešetřil and Rödl (1975a).

Theorem 5.6. *For every $k > 1$, the class* Forb(K_k) *has the A-Ramsey property iff A is either discrete or a complete graph with $< k$ vertices.*

In one direction (necessity) one uses an appropriate version of the Ordering lemma. The sufficiency is harder of course. We shall discuss the relevant proof methods in section 7.

But the development has been rapid and it follows from Rödl and Ruciński (1995) that for any choice k, t, almost all graphs G_n with n vertices which do not contain K_{k+1} are Ramsey for K_k: $G_n \rightarrow (K_k)_t^2$.

In view of the previous section it is expected that one can prove a more general statement related to subclasses of Soc(Δ). We shall do so by means of the following definitions.

A structure $A = (X, \mathcal{M})$, $\mathcal{M} = (\mathcal{M}_\delta; \delta \in \Delta)$, (of type Δ), is said to be *irreducible* if for every pair x, $y \in X$, there exists $\delta \in \Delta$ and $M \in \mathcal{M}_\delta$ such that x, $y \in M$.

Let $\mathcal{F}$ be a (possibly infinite) set of structures (of type Δ). Denote by Forb$_\Delta(\mathcal{F})$ the class of all structures A (of type Δ) which do not contain any member of $\mathcal{F}$ as a substructure.

Now we can formulate the principal result for set structures due to Nešetřil and Rödl (1977b, 1983a).

Theorem 5.7 (Ramsey classes of structures). *Let Δ be a type. Let $\mathcal{F}$ be a (possibly infinite) set of irreducible structures (of type Δ). Then* Forb$_\Delta(\mathcal{F})$ *is a Ramsey class.*

This is a generalization of the previous theorems. For example, by taking $\Delta = \{2\}$, $\mathcal{F} = \{K_k\}$ we get (a nontrivial) generalization of Theorem 5.1 to partitions of ordered subgraphs.

As we shall see now, this is in fact as far as we can go on this level of generality. Let us be more specific and let us analyse the Ramsey classes of graphs in more detail.

First, in discussing Ramsey classes we can restrict ourselves to hereditary classes. For nonhereditary classes we have in general no hope for a characterization: any sequence G_1, G_2, G_3 . . . with $G_{i+1} \rightarrow (G_i)_2^{G_{i-1}}$ will form a Ramsey class. Also in view of the Ordering lemma we have to deal with ordered graphs.

If K is a Ramsey class (of structures of type Δ) and we permute the colors of structural colorings we get a Ramsey class again. Particularly, if K is a Ramsey class of graphs and $\bar{K}$ denotes the class of all complements of graphs from K, then $\bar{K}$ is again a Ramsey class of graphs.

If K and K' are hereditary Ramsey classes then the class $K \cup K'$ of all structures which belong either to K or to K' is clearly a Ramsey class again.

We have seen that the class Compl is a Ramsey class. By a simple "product" argument it follows that the class Eq of all equivalences (i.e., of disjoint unions of complete graphs with a linearly ordered set of vertices) is a Ramsey class. By the above remark also the class $\overline{\mathrm{Eq}}$ is also a Ramsey class of graphs. The class $\overline{\mathrm{Eq}}$ consists of all complete multipartite graphs, i.e., *Turán graphs*.

A bit surprisingly, these remarks and the previously stated theorems exhaust all Ramsey classes of graphs.

Theorem 5.8 (Characterization of Ramsey classes of graphs). *For a hereditary class K of graphs the following two statements are equivalent*:

(1) *The class of all ordered graphs from K is a Ramsey class of graphs*;

(2) *K is a union of classes K_i; $i \in I$, each of which is either a class* $\mathrm{Forb}(K_k)$ or $\overline{\mathrm{Forb}(K_k)}$ (*i.e., the class* $\{\bar{G}; G \in \mathrm{Forb}(K_k)\}$), *or the class* Eq, *or the class* $\overline{\mathrm{Eq}}$.

This result was proved in Nešetřil (1989). The proof uses a deep result of Lachlan and Woodrow (1980) on graph-amalgamations.

For types Δ of different form, i.e., $\Delta \neq \{2\}$ (particularly for $\Delta = \{3\}$) a characterization of Ramsey classes of structures is not known. On the other hand no infinite Ramsey class different from the $\mathrm{Forb}_\Delta(\mathcal{F})$ above is known. Perhaps even no essentially different Ramsey classes exist.

Various simplified proofs of the main result of this section are known, see, e.g., Nešetřil and Rödl (1981, 1982, 1987b). These results are mostly based on the *amalgamation technique*, known as "partite construction", which is due to Nešetřil and Rödl (1981, 1982). We give a sample of this method in section 7. Also, restricted Ramsey-type theorems were extended to all other Ramsey classes of structures, such as parameter sets or spaces. This has been done in Frankl et al. (1987), Nešetřil and Rödl (1987b) and Prömel and Voigt (1988).

In section 7 we are going to prove a structural Ramsey theorem (Theorem 5.3) and its forbidden version (Theorem 5.7).

6. The Erdős–Ramsey problem

What is the structure of graphs G which arrow the triangle (i.e., which graphs G satisfy $G \to (K_3)_t^{K_2}$ for a fixed t)?

Which classes of graphs have the edge-Ramsey property?

Which graphs have to be contained in every graph G which arrows the triangle?

For which sets $\mathcal{F}$ does the class $\mathrm{Forb}(\mathcal{F})$ have the edge-Ramsey property?

These problems and their analogues for other classes may be regarded as the central questions of structural extensions of Ramsey's theorem. Viewing the difficulties which one encounters even when proving the basic Ramsey-type theorems, these problems may be regarded as too ambitious. In fact, they were never formulated (at least explicitly) on this level of generality since one could not solve even particular cases (such as the above triangle case). However, special cases have been asked by Erdős, Hajnal, Galvin in the late 60s. During the development of the field, many particular questions were related to the above problems. We propose to call the above problems *Erdős–Ramsey problems* as Erdős' persistence has been largely responsible for the development of this area.

In order to formulate and explain our approach it appears that we have to investigate Ramsey properties in the context of chromatic number.

6.1. Ramsey properties via chromatic number

In a similar way as one can view most extremal problems as questions concerned with the independence number of a (*special*) hypergraph one can relate Ramsey-type theorems to the chromatic number of a (special) hypergraph. This can be done in full generality for a class K endowed with subobjects (as introduced in the previous section) by means of the following construct.

Given three objects A, B, C of K denote by $\langle A, B, C\rangle$ the hypergraph $(X, \mathcal{M})$ defined as follows:

$$X = \binom{C}{A},$$

$$\mathcal{M} = \left\{ \binom{B'}{A} : B' \in \binom{C}{B} \right\}.$$

$\langle A, B, C\rangle$ is sometimes called the *triangle-hypergraph* since it arises by considering the scheme

$$\begin{tikzcd} B \arrow[r] & C \\ A \arrow[u] \arrow[ur] & \end{tikzcd} \tag{6.1}$$

Recall that the chromatic number $\chi(H)$ of a hypergraph H is the minimum number of colors which suffice for a coloring of vertices in such a way that no monochromatic edge occurs, and the independence number $\alpha(H)$ is the maximum number of vertices which do not contain any edge.

If we compare the corresponding definitions then we immediately obtain the following statement valid for *every* class K.

Proposition 6.1. *For every choice of objects A, B, C, of K and every positive integer t we have*

$$C \to (B)^A_t \quad \textit{iff} \quad \chi(\langle A, B, C\rangle) > t.$$

This observation may serve as one of the possible approaches to Ramsey-type statements.

This formulation may be convenient in other contexts as well. Let us give some examples.

(1) We say that an *A-density theorem* holds (in a class K) if for every $\varepsilon > 0$ and every $B \in K$ there exists $C \in K$ with

$$\alpha(\langle A, B, C\rangle) < \varepsilon \left|\binom{C}{A}\right| .$$

The validity of an A-density theorem *always* implies the A-Ramsey property since

$$\chi(\langle A, B, C\rangle) \geqslant \frac{|\binom{C}{A}|}{\alpha(\langle A, B, C\rangle)} .$$

(2) The above reformulations of Ramsey and extremal (Turán) problems indicate that one could study these problems in a unified way. In fact, this has been done and so-called *Ramsey–Turán problems* were studied in a series of papers by Erdős, Sós and others, see, e.g., Erdős et al. (1993) and Erdős and Sós (1982).

(3) One can study the *quantitative* form of Ramsey-type theorems: Given a coloring $c: \binom{C}{A} \to [t]$ denote by $\binom{C}{B}_c$ the size of the subset of $\binom{C}{B}$ formed by all c-homogeneous copies of B. Thus $C \to (B)_t^A$ iff $\binom{C}{B}_c > 0$ for every coloring c.

It appears that much more is true: all known Ramsey classes contain objects for which

$$\binom{C}{B}_c > \varepsilon \left|\binom{C}{B}\right| , \tag{6.2}$$

where ε is a positive constant depending on A, B, t (and independent of C). Moreover, we know that for all full Ramsey theorems, every sufficiently large object has property (6.2).

Such results were obtained in a series of papers by Frankl et al. (1988, 1989).

(The basic idea can be illustrated on Ramsey's theorem itself: Given n, p, t, set $r = r(p, t, n)$ and consider $R > r$. Then obviously

$$\binom{R}{n}_c \geqslant \binom{R}{r} \Big/ \binom{R-n}{r-n} = \binom{r}{n}^{-1} \cdot \binom{R}{n} .)$$

(4) As we have seen, Ramsey's theorem corresponds to the chromatic number of the triangle hypergraph $\langle A, B, C\rangle$ which in turn, corresponds to diagram (6.1). The dual Ramsey theorem then corresponds to the following diagram:

$$\begin{tikzcd} B \arrow[d] & C \arrow[l] \arrow[dl] \\ A & \end{tikzcd} \tag{6.3}$$

Let us return to our main theme. There is no hope (at least at present) to use

Proposition 6.1. for a proof of a Ramsey-type statement which is not related to partitions of singletons. The difficulty lies in the lack of techniques for "forcing" a large chromatic number: The triple hypergraphs $\langle A, B, C\rangle$ have many special properties which are difficult (and, in full generality, probably impossible) to control.

However, the triangle hypergraphs $\langle A, B, C\rangle$ may be used rather efficiently for an analysis of the necessary conditions which Ramsey objects have to satisfy.

In fact this approach appears to be the unique tool for characterization of natural local obstacles for Ramsey objects. This is based on the following.

Folklore lemma 6.2. *Let $H = (V, E)$ be a k-uniform hypergraph with an infinite chromatic number. Then H contains every finite k-tree.*

An example of a k-tree is depicted in fig. 6.1 and we may also define k-trees recursively by means of amalgamations. A set S of t edges which pairwise intersect in a single vertex x is called a *t-star* S; x is the center of S.

Proof. We start with the following.

Claim. *Let H be a hypergraph with chromatic number $t + 1$. Then H contains a t-star.*

Proof of Claim. Put $H = (V, E)$. Let $V = V_1 \cup \cdots \cup V_{t+1}$ be a coloring such that V_1 is a maximal independent set in $H - V_1$, etc. It follows that for every $v \in V_{t+1}$, every set $V_1 \cup \{v\}, V_2 \cup \{v\}, \ldots, V_t \cup \{v\}$ contains an edge of H. These sets form a t-star centered at v. □

Using the claim, we prove the Folklore lemma quite easily. Fix a finite k-tree T with n vertices. Set $t = n^2$. We prove that every finite hypergraph $H = (V, E)$ with $\chi(H) > t$ contains T. Define recursively subsets $V_0, V_1, \ldots, V_n$ of V as follows: Set $V_0 = V$ and let V_{i+1} be the set of all vertices of the hypergraph $H_i = H|V_i$ which are centers of an n-star in H_i. It follows from the claim that $\chi(H_1) > t - n$ and, generally, $\chi(H_i) > t - in$. Consequently, $V_n \neq \emptyset$. Fix $x_n \in V_n$ and let the edges of $H_{n-1} e_n^{n-1}, \ldots, e_n^{n-1}$ form an n-star at x_n. Each vertex x of $\bigcup_{i=1}^{n} e_i^{n-1}$ is the

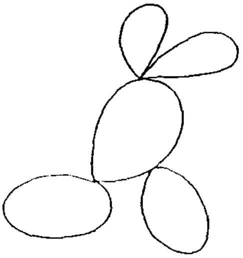


Figure 6.1.

center of an n-star in H_{n-2}, etc. Some of these edges may then be used for a construction of a tree isomorphic to T. □

The importance of the Folklore lemma is that it characterizes all finite unavoidable subgraphs of highly chromatic graphs.

Theorem 6.3. *For a finite k-uniform hypergraph F, the following two statements are equivalent*:

(1) *F is a subgraph of every k-uniform hypergraph H with infinite chromatic number*;

(2) *F is a k-tree.*

(In (1) one can assume that F is a subgraph of every k-uniform hypergraph with a sufficiently large finite chromatic number.)

Theorem 6.3 is a profound result. Its nontrivial part $(1) \Rightarrow (2)$ asserts that for every choice of positive integers t and l, there exists a k-uniform hypergraph H with the following properties:

(i) $\chi(H) > t$,

(ii) $g(H) > l$,

where $g(H)$ denotes the girth of H (alternatively, this means that H does not contain cycles of length $\leqslant l$).

This is a combinatorial classic which started in the 40s with Tutte (1954) and Zykov (1949) for the case $k = 2$, $l = 3$. The general case $k = 2$ was solved by Erdős (1959) in his seminal paper by a striking application of the probabilistic method. The same method has been modified in Erdős and Hajnal (1966) to yield the general result.

For many reasons it is desirable to have a constructive proof of Theorem 6.3. This appeared to be difficult and a construction in full generality was finally given by Lovász (1968). A simplified construction has been found in the context of Ramsey theory by Nešetřil and Rödl (1979a). The graphs and hypergraphs with the above properties (i), (ii) are called *highly chromatic* (*locally*) *sparse graphs*, for short.

Their existence could be regarded as one of the true paradoxes of finite set theory and it has always been felt that this result is one of the central results in combinatorics.

Recently it has been realized that sparse and complex graphs may be used in theoretical computer science for the design of fast algorithms. However, what is needed there is not only a construction of these "paradoxical" structures but also their reasonable size. In one of the most striking recent developments, a program for constructing complex sparse graphs has been successfully carried out. Using several highly ingenious constructions which combine algebraic and topological methods it has been shown that there are complex sparse graphs, the size of which in several instances improves the size of random objects. See Margulis (1975), Alon (1986), Lubotzky et al. (1988) and chapter 32.

Particularly, it follows from Lubotzky et al. (1988) that there are examples of graphs with girth l, chromatic number t and the size at most t^{3l} (the Erdős probabilistic bound being t^l). Prior to this construction, not even a primitive recursive upper bound was known. But not everything in this direction is presently known. Below we shall see that for Ramsey structures this is still an open problem. Also, a bit surprisingly, the following is still open.

Problem. Find a primitive recursive construction of highly chromatic locally sparse p-uniform hypergraphs. Indeed, even triple systems (i.e., $p = 3$) present a problem.

Another implication of Theorem 6.3 is the following result (Nešetřil and Rödl 1966).

Corollary 6.4. *Let A be a finite set of 2-connected graphs. Then the class* Forb(A) *has the vertex-partition property.*

As indicated above (see remarks before Theorem 5.8) if the class K has the vertex-partition property then the class $-K$ (formed by complements of graphs from K) and the class $+K$ (formed by disjoint unions of graphs from K) have the vertex-partition property as well. These constructions may be iterated (to yield classes $+-+-+-K$). Then Corollary 6.4 together with the union of the classes $+-\cdots+-$ (Compl.) are the only known classes of the form Forb(A) which have the vertex-partition property. The problem is not yet completely solved, see Rödl and Sauer (1992), and Rödl et al. (1995).

6.2. *Ramsey graphs*

One can view Theorem 6.3 as the solution of the Ramsey problem for partitions of singletons (of set systems). For other Ramsey type questions, the situation is less transparent and progress has been slow. Since this is not the distant past in which theorems melt into definitions, history is vital and we should review it. Of course, the Folklore lemma 6.2 gives us a good candidate for the solution of the Erdős–Ramsey problem. But this was not the real trajectory of discovery. It took some time before the pattern of the problem was recognized.

Let us illustrate the main theorem of this chapter again by the particular example of edge-Ramsey graphs for triangles (compare the introduction to section 5).

Thus, let us study those graphs G which have the edge-Ramsey property for the triangle K_3; explicitly, we study G with $G \longrightarrow (K_3)^2_t$. Since we are interested in the structural properties of G, we have to assume that t is sufficiently large (to avoid singular small examples). By the restricted Ramsey theorem 5.5 we know that G may be chosen K_4-free and that the class of all K_4-free graphs is in fact a Ramsey class (by Theorem 5.7). According to Theorem 5.8 we know that on this level of generality we have no other results. However, we are interested in the edge-colorings only and thus we can hope for stronger results.

The first step was taken by Spencer (1975b). He showed that for every choice of positive integers t and l, there exists a graph G and a system $\mathscr{T} \subseteq \binom{E}{3}$ such that:

(i) every $T \in \mathscr{T}$ corresponds to a triangle in G;

(ii) the hypergraph $(E, \mathscr{T})$ has no cycles of length $< l$;

(iii) $\chi(E, \mathscr{T}) > t$ and thus $\mathscr{T}$ is a Ramsey family of triangles in G.

We call this (and the like) result, a *sparse Ramsey family* (of copies of A in B).

One can prove the existence of sparse Ramsey families for every instance of the full Ramsey theorem since we may use probabilistic means (in the spirit of Erdős' 1959 original ideas; see Spencer 1975b for details).

However, this does not imply restricted theorems since one constructs $\mathscr{T}$ for essentially a complete graph G. The situation is less clear and more in the spirit of restricted theorems if we demand in the above result that $\mathscr{T}$ is the triple system induced by *all* triangles in G and yet the conditions (ii) and (iii) above are satisfied. This has been posed as a problem in Spencer (1975b) and solved affirmatively in Nešetřil and Rödl (1984, 1989). Again one can prove the analogous statement for every class with a full Ramsey theorem. However, until recently this presented a problem for structures different from set systems, especially for parameter sets and spaces. Recently, however, all of these problems were resolved. We wish to single out the following result which has been obtained by Rödl (1986a) and Nešetřil and Rödl (1987b).

Theorem 6.5 (Sparse Hales–Jewett theorem). *Let t, l be positive integers, A a non-empty set. Then there exists an N and a set $X \subseteq A^n$ with the following properties*:

(1) *For every t-coloring of X there exists a monochromatic line in X; we denote these by $X \to (1)^0_t$.*

(2) *All lines in X (i.e., $\binom{X}{1}$) do not form cycles of length $< l$; explicitly the hypergraph $(X, \binom{X}{1})$ has girth $\geqslant l$.*

Actually, this is a key result from which one can derive other results of this type quite easily, see Nešetřil and Rödl (1989, 1990b), and compare also section 7.

One is tempted to call these and the like results sparse Ramsey-type theorems. However, this is misleading. A more appropriate name would perhaps be *Ramsey-type theorems for sparse copies*. For nonsingleton partitions there is still a long way to go.

Let us illustrate this again on our example of a Ramsey graph G for the triangle K_3.

A Ramsey theorem for sparse copies of K_3 claims that we can find G with $G \to (K_3)^2_t$ and yet G does not contain pictures like those depicted in fig. 6.2. These graphs reflect examples of cycles in *edges*.

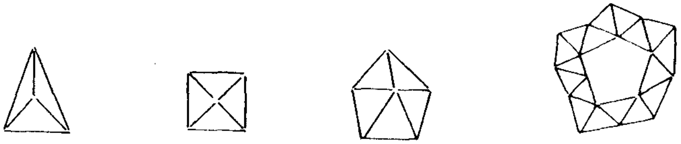


Figure 6.2.

No information is available about subgraphs whose cycle structure is not induced by edges such as those depicted in fig. 6.3.

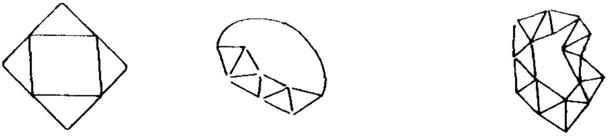


Figure 6.3.

The situation is a bit confusing since every graph G with $G \rightarrow (K_3)_2^2$ has to contain a C_4. However, there is no reason why it should contain a *chordless* C_4.

In this way it became apparent at the end of the 70s (see, e.g., Nešetřil and Rödl 1979b) that the following peculiar problem plays a central role: Does there exist for every bipartite graph G, a bipartite graph H with the following properties:

$$\begin{aligned} &\text{(i)} \quad H \rightarrow (G)_2^2\,; \\ &\text{(ii)} \quad H \text{ has the same girth as } G. \end{aligned} \tag{6.4}$$

However, (6.4) appeared to be just as difficult as the Erdős–Ramsey problem and, in fact, Erdős (1975), motivated by the strict finiteness of this problem conjectured that the answer is negative already for girth 5 (i.e., for rectangle-free bipartite graphs).

This particular case (i.e., of girth 5) of (6.4) has been solved by Nešetřil and Rödl in (1987a). Their method does not generalize to girth >6. However, it is related to the following result which is of independent interest.

Theorem 6.6 (Ramsey theorem for simple systems). *Let $\mathscr{L}_k$ denote the class of all k-uniform hypergraphs which are simple (i.e., every two edges meet in at most one point). Then the distinct class $\mathscr{L}_k$ has the edge-Ramsey property.*

Explicitly: for every $B \in \mathscr{L}_k$, there exists $C \in \mathscr{L}_k$ such that $C \rightarrow (B)_t^e$ where e stands for the single-edge hypergraph.

By a theorem of Wilson (see chapter 14) this leads to the following.

Corollary 6.7 (Ramsey theorem for Steiner systems). *Let $\mathscr{S}_k$ denote the class of all Steiner k-systems. Then $\mathscr{S}_k$ has the edge-Ramsey property.*

(One can easily see that Corollary 6.7 does not hold for block designs with $\lambda \neq 1$.)

These results were essentially the first results which went beyond the classes Forb($\mathscr{F}$). Only very recently was the above problem (6.4) solved affirmatively by

Nešetřil and Rödl. This result is not yet published. As indicated above, in particular, this yields a solution of the Erdős–Ramsey problem for triangles.

7. Techniques

Ramsey theory benefits from most combinatorial (and fortunately also non-combinatorial) methods. Perhaps the very fact that many different tools were applied in order to solve particular problems essentially contributed to the popularity of the field. This diversity is well documented, e.g., by Erdős et al. (1984a), Graham (1981), Graham et al. (1980), Grinstead and Roberts (1982) and Nešetřil and Rödl (1979b), and most recently by a collection of papers in Nešetřil and Rödl (1990a).

In this paper we followed the mainstream of Ramsey theory formed by efforts to prove the strongest structural Ramsey-type results. Even in this limited scope we had to omit several interesting and very active areas. Some of them were mentioned earlier. Let us complement this by mentioning that we left out structural extensions of the canonical lemma of Erdős and Rado (1950). These extensions were studied in great detail by Prömel and Voigt; see, e.g., Prömel and Voigt (1985). Also we had to omit various applications of Ramsey theory.

Let us finish this paper by giving a sample of the proof technique which proved to be quite useful in this area. This is the *amalgamation technique* due to Nešetřil and Rödl.

The method originated in 1976 from the analysis of methods of Nešetřil and Rödl (1977b), e.g., Nešetřil and Rödl 1981, 1982, 1987b for examples of its use).

We shall illustrate this by giving a short proof of the structural Ramsey theorem 5.3.

It is an important feature of the amalgamation technique that it is self-refining and consequently we can derive a number of corollaries such as Theorems 5.7 and 6.5.

The amalgamation technique consists of three basic steps:

(i) Definition of partite systems and their amalgamation.

(ii) The partite lemma.

(iii) The partite construction.

All versions of the amalgamation technique have this structure. The following proof of the structural Ramsey theorem and Ramsey classes of structures follows Nešetřil and Rödl (1989).

7.1. *Partite systems* (*of type* Δ)

An *a-partite system* A is a pair $((X_i)_{i=1}^{a}, \mathcal{M})$ where

(a) $X = \bigcup_{i=1}^{a} X_i$ is an ordered set satisfying $X_1 < X_2 < \cdots < X_a$,

(b) $\mathcal{M} = (\mathcal{M}_\delta; \delta \in \Delta)$, $\mathcal{M}_\delta \subseteq \binom{X}{n_\delta}$,

(c) $|M \cap X_i| \leq 1$ for every $M \in \mathcal{M}_\delta$, $i = 1, \ldots, a$, $\delta \in \Delta$.

The sets X_i are called *parts* of A, and elements of $\mathcal{M}$ are called edges of A. Property (c) implies that edges are transversals with respect to the family

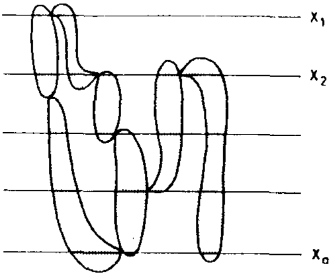


Figure 7.1.

$X_1 < \cdots < X_a$. Given a subset $Y \subseteq X$, we denote by $\mathrm{tr}(Y)$ the *trace* of Y, i.e., the set $\{i\colon Y \cap X_i \neq \emptyset\}$. See fig. 7.1.

$\boldsymbol{A}$ is called a *transversal* if $|X_i| = 1$ for every $i = 1, \ldots, a$. $\boldsymbol{A}$ is a *subsystem* of $\boldsymbol{B} = ((Y_i)_{i=1}^{b}, \mathcal{N})$ if there exists a monotone injection $i\colon \{1, \ldots, a\} \to \{1, \ldots, b\}$ such that $X_i \subseteq Y_i$ for $i = 1, \ldots, a$ and $\mathcal{M}_\delta = \mathcal{N}_\delta \cap \binom{X}{n_\delta}$ for $\delta \in \Delta$. An *isomorphism* is defined as an order- and parts-preserving isomorphism. A subsystem of $\boldsymbol{B}$, isomorphic to $\boldsymbol{A}$ is called a *copy* of $\boldsymbol{A}$ in $\boldsymbol{B}$, the set of such copies is denoted (again) by $\binom{\boldsymbol{B}}{\boldsymbol{A}}$.

7.2. *The partite lemma*

Lemma 7.1. *Let t be a positive integer and let $\boldsymbol{A}$ and $\boldsymbol{B}$ be a-partite systems. Moreover, let $\boldsymbol{A}$ be a transversal. Then there exists an a-partite system $\boldsymbol{C}$ such that*

$$\boldsymbol{C} \to (\boldsymbol{B})_t^{\boldsymbol{A}}\ .$$

Here the arrow notation has the same meaning as the one above (with $\binom{\boldsymbol{C}}{\boldsymbol{B}}$ being the set of all copies of $\boldsymbol{B}$ in $\boldsymbol{C}$, i.e., as a-partite subsystems).

Proof. Set $\boldsymbol{A} = ((X_i)_{i=1}^{a}, \mathcal{M})$, $\boldsymbol{B} = ((Y_i)_{i=1}^{a}, \mathcal{N})$. Since $\boldsymbol{A}$ is a transversal we may suppose without loss of generality that $\bigcup_{\delta \in \Delta} \mathcal{M}_\delta$ is the set of all subsets of X (this may be achieved by adding "dummy" edges to $\mathcal{M}$ and $\mathcal{N}$). Without loss of generality we may also assume that every vertex $y \in Y$ is contained in a copy of $\boldsymbol{A}$; this is a general comment: if $\boldsymbol{B}^*$ is the subsystem of $\boldsymbol{B}$ induced by $\binom{\boldsymbol{B}}{\boldsymbol{A}}$ and $\boldsymbol{C}^* \to (\boldsymbol{B}^*)_t^{\boldsymbol{A}}$ then we may easily construct $\boldsymbol{C}$ such that $\boldsymbol{C} \to (\boldsymbol{B})_t^{\boldsymbol{A}}$ by enlarging every $\boldsymbol{B}^* \in \binom{\boldsymbol{C}^*}{\boldsymbol{B}^*}$ to a system $\boldsymbol{B}$.

Now take N to be sufficiently large (indeed, very large) number. Define an a-partite system $\boldsymbol{C} = ((Z_i)_{i=1}^{a}, \mathcal{O})$, $\mathcal{O} = (\mathcal{O}_\delta; \delta \in \Delta)$ as follows: $Z_i = Y_i \times \cdots \times Y_i$ (N times); i.e., each element of Z_i has the form $(x_j\colon x_j \in Y_i, j = 1, \ldots, N)$.

Set $Z = \bigcup_{i=1}^{a} Z_i$ and for $j = 1, \ldots, N$, define the projection $\pi_j\colon Z \to Y$ by $\pi_j(x_k\colon k = 1, \ldots, N) = x_j$.

Clearly π_j maps Z_k into Y_k.

We define $\mathcal{O} = (\mathcal{O}_\delta; \delta \in \Delta)$ as follows: First put $\mathcal{N}_\delta = \mathcal{N}'_\delta \cup \mathcal{N}''_\delta$ where $\mathcal{N}'_\delta$ is the set of all edges of n which belong to a copy of $\boldsymbol{A}$ in $\boldsymbol{B}$, $\mathcal{N}''_\delta = \mathcal{N}_\delta - \mathcal{N}'_\delta$ (note that in general we cannot assume $\mathcal{N}''_\delta = \emptyset$). We put

$$\{(x_j^k: j = 1, \ldots, N): k = 1, \ldots, n_\delta\} \in \mathcal{O}_\delta$$

if $\operatorname{tr}(\{x_j^k: k = 1, \ldots, n_\delta\}) = \operatorname{tr}(\{x_{j'}^k: k = 1, \ldots, n_\delta\})$ for all j, $j' \leqslant N$, and one of the following possibilities occurs:

(1) $\{x_j^k: k = 1, \ldots, n_\delta\} \in \mathcal{N}'_\delta$ for every $j = 1, \ldots, N$.

(2) There exists a non-empty set $\Lambda \subseteq \{1, \ldots, N\}$ such that

$$\{x_j^k: k = 1, \ldots, n_\delta\} = \{x_{j'}^k: k = 1, \ldots, n_\delta\} \in \mathcal{N}''_\delta$$

for all j, $j' \in \Lambda$ and

$$\{x_j^k: k = 1, \ldots, n_\delta\} \in \mathcal{N}'_\eta \quad \text{for all } j \notin \Lambda\,.$$

(Note that, in general, $\eta \neq \delta$; however, η is uniquely determined by $\operatorname{tr}(\{x_j^k: k = 1, \ldots, n_\delta\})$.

We shall prove $\boldsymbol{C} \to (\boldsymbol{B})_t^{\boldsymbol{A}}$ provided N is large enough. This easily follows from the two facts stated below as Fact 1 and 2.

Fact 1. *$A' \in \binom{C}{A}$ iff $\pi_j(A') \in \binom{B}{A}$ for every $j = 1, \ldots, N$.*

Proof. Obvious by the definition of $\mathcal{O}$. □

Set $\binom{B}{A} = \{\boldsymbol{A}_1, \ldots, \boldsymbol{A}_r\}$. Set $R = \{1, \ldots, r\}$. Think of the set R^N endowed with Hales–Jewett (combinatorial) lines. A *line* is a set L of the following form: Fix $\Lambda \subseteq \{1, \ldots, N\}$ (non-empty) and $\alpha^0 = (\alpha_1^0, \ldots, \alpha_N^0) \in R^N$ and set

$$L = \{(\alpha_1, \ldots, \alpha_N): \alpha_i = \alpha_i^0 \text{ for } i \notin \Lambda, \alpha_i = \alpha_j \text{ for } i, j \in \Lambda\}\,.$$

Clearly $|L| = r$. Given $\alpha = (\alpha_1, \ldots, \alpha_N) \in R^N$, denote by $V(\alpha)$ the set of all vertices x of Z which satisfy $\pi_j(x) \in A_{\alpha j}$. Set $V(L) = \bigcup_{\alpha \in L_N} V(\alpha)$. By virtue of Fact 1, the set $\binom{C}{A}$ is in 1–1 correspondence with the set R^N.

Fact 2. *Let L be a line of R^N. Then $V(L)$ induces a copy of $\boldsymbol{B}$ in $\boldsymbol{C}$.*

Proof. Check the definition of C. □

Now we invoke the Hales–Jewett theorem 2.2 and choose N sufficiently large so that for every partition of R^N into t classes, one of the classes contains a monochromatic line. This implies $\boldsymbol{C} \to (\boldsymbol{B})_t^{\boldsymbol{A}}$. Indeed, let $\binom{C}{A} = \mathcal{A}_1 \cup \cdots \cup \mathcal{A}_t$ be a partition. By Fact 1 this induces a partition $R^N = \mathcal{A}'_1 \cup \cdots \cup \mathcal{A}'_t$, by defining $a \in \mathcal{A}'_i$ iff $V(\alpha)$ induces a copy belonging to $\mathcal{A}_i$. By the Hales–Jewett theorem there exists a monochromatic line L in R^N. This in turn, using Fact 2, yields $\boldsymbol{B}' \in \binom{C}{B}$ such that $\binom{B'}{A}$ is monochromatic. □

7.3. The partite construction

In this part we prove the structural Ramsey theorem by means of the partite construction. We follow closely the construction given in Nešetřil and Rödl (1981, 1982). Let k, A, B be fixed. Consider $\boldsymbol{A}$ as a transversal a-partite system and $\boldsymbol{B}$ as a transversal b-partite system. Set explicitly

$$\boldsymbol{B} = (\{y_1, \ldots, y_b\}, \mathcal{M}) .$$

Set $p = r(a, t, b) = \min\{n : n \to (b)^a_t\}$ (the classical Ramsey number). Set $q = \binom{p}{b}$, $\binom{\{1, \ldots, p\}}{a} = \{M^1, \ldots, M^q\}$. We shall construct "pictures" $\boldsymbol{P}^0, \ldots, \boldsymbol{P}^k, \ldots, \boldsymbol{P}^q$ by induction on k. Picture $\boldsymbol{P}^q$ will be the desired system $\boldsymbol{C}$.

Let $\boldsymbol{P}^0 = ((X^0_i)^p_{i=1}, \mathcal{O})$ be a p-partite system where for each choice of b parts $X^0_{i_1}, \ldots, X^0_{i_b}$, the subsystem of $\boldsymbol{P}^0$ induced by them contains a copy of $\boldsymbol{B}$. Such a "picture" $\boldsymbol{P}^0$ may be formed as a disjoint union of copies of $\boldsymbol{B}$.

Suppose picture $\boldsymbol{P}^k = ((X^k_i)^p_{i=1}, \mathcal{O}^k)$, $k < q$, is given. Consider M^{k+1} and the a-partite system $\boldsymbol{D}^{k+1}$ induced in $\boldsymbol{P}^k$ by parts X^k_i where each y_i belongs to M^{k+1}. By the partite lemma there exists an a-partite system $\boldsymbol{E}^{k+1}$ such that

$$\boldsymbol{E}^{k+1} \to (\boldsymbol{D}^{k+1})^A_t .$$

Extend each copy of $\boldsymbol{D}^{k+1}$ in $\boldsymbol{E}^{k+1}$ to a copy of $\boldsymbol{P}^k$ in such a way that the distinct copies of $\boldsymbol{P}^k$ intersect only in vertices of $\boldsymbol{E}^{k+1}$. In this amalgamation the parts of distinct copies of $\boldsymbol{P}^k$ are preserved.

Denote the resulting amalgamation by $\binom{\boldsymbol{E}^{k+1}}{\boldsymbol{D}^{k+1}} * \boldsymbol{P}^k$. (For a more explicit definition, see Nešetřil and Rödl (1984). Set $\boldsymbol{P}^{k+1} = \binom{\boldsymbol{E}^{k+1}}{\boldsymbol{D}^{k+1}} * \boldsymbol{P}^k$ (see fig. 7.2). Finally, put $\boldsymbol{C} = \boldsymbol{P}^q$. We claim that $\boldsymbol{C}$ has the desired properties.

Claim 1. *Every irreducible subsystem in* $\boldsymbol{C}$ *is a subsystem of* $\boldsymbol{B}$.

Proof. Induction on k. This being trivial for $k = 0$, in the inductive step the amalgamation does not create any new irreducible system. □

Claim 2. $\boldsymbol{C} \to (\boldsymbol{B})^A_t$.

Proof. Backward induction on $k = q, q-1, \ldots, 1$. In the inductive step $(k+1) \to k$ we apply the partite lemma and find a copy of $\boldsymbol{P}^k$ such that all copies of $\boldsymbol{A}$ with trace M^k are monochromatic. This leaves us, for $k = 0$, with a copy $\boldsymbol{P}'$ of $\boldsymbol{P}^0$

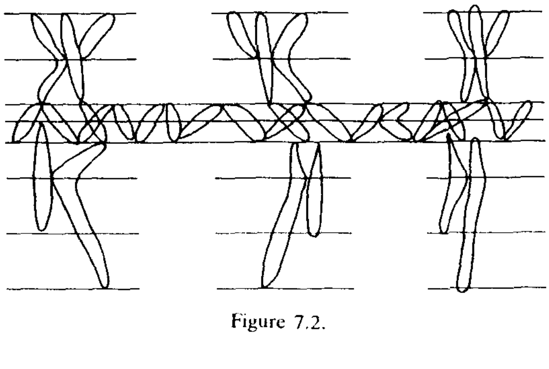


Figure 7.2.

in C where the color of a copy of A in P depends only on its trace. However, such a copy of P^0 contains a monochromatic copy of B by the construction of P^0 and the assumption that $p \rightarrow (b)^a_t$. □

Proof of Theorem 6.5 (Ramsey theorem for simple hypergraphs). Note that transversal edges intersect in at most one vertex and that lines form a simple hypergraph. Consequently, for two different lines L and L', the sets $V(L)$ and $V(L')$ intersect either in a single vertex or a single edge (see the above proof of the partite lemma). □

Only recently has it been discovered that one can apply the amalgamation technique to "algebraic" Ramsey-type theorems (such as the Van der Waerden and vector space theorem). This has been done in Frankl et al. (1987), Nešetřil and Rödl (1987b) and Prömel and Voigt (1988). Particularly, Frankl et al. (1987) contains an induced and restricted space theorem, (i.e., the analogues of Theorems 5.1 and 5.5 for spaces) and Nešetřil and Rödl (1987b) contains a structural space theorem and a Ramsey classes of structures-theorem (i.e., analogues of Theorems 5.4 and 5.7 for spaces). Let us finally remark that once a proper amalgamation pattern has been realized one can proceed in complete analogy with the methods described in this section. To see this one should compare the papers Nešetřil and Rödl (1989, 1990b).

References

Abramson, F.G., and L.A. Harrington
[1978] Models without indiscernibles, *J. Symbolic Logic* **43**, 572–600.

Ajtai, M., J. Komlós and E. Szemerédi
[1980] A note on Ramsey numbers, *J. Combin. Theory A* **29**, 354–360.
[1981a] A dense infinite Sidon sequence, *European J. Combin.* **2**, 1–15.
[1981b] On Turán's theorem for sparse graphs, *Combinatorica* **1**, 313–317.

Alon, N.
[1986] Eigenvalues, geometric expanders, sorting in sounds, and Ramsey theory, *Combinatorica G.* **3**, 207–219.
[1990] Ramsey graphs cannot be defined by polynomials, *J. Graph Theory* **14**, 651–661.
[1991] Non-constructive proofs in combinatorics, in: *Proc. ICM, 1990* (Springer, Berlin) pp. 1421–1429.
[1994] Explicit Ramsey graphs and orthonormal labellings, *Electron. J. Combin.* 1, R12 (8pp).

Alon, N., and J. Spencer
[1992] *Probabilistic Method* (Wiley, New York).

Alon, N., H. Lefmann and V. Rödl
[1991] On an anti-Ramsey type result, *Colloq. Math. Soc. János Bolyai* **60**, 9–22.

Alon, N., R.A. Duke, H. Lefmann, V. Rödl and R. Yuster
[1994] The algorithmic aspects of the regularity lemma, *J. Algorithms* **16**, 80–109.

Alon, N., I. Kříž and J. Nešetřil
[1995] How to color shifted hypergraphs, *Studia Sci. Math. Hungar.*, to appear.

Babai, L.
[1985] An anti-Ramsey theorem, *Graphs and Combin.* **1**, 23–28.

Babai, L., and V.T. Sós
[1985] Sidon sets in groups and induced subgraphs of Cayley graphs, *European J. Combin.* **6**, 101–114.

Beck, J.

[1991] An algorithmic approach to the Lovász Local Lemma I, *Random Structures and Algorithms* **2**, 343–365.

Behrend, F.A.

[1946] On sets of integers which contain no three in arithmetic progression, *Proc. Nat. Acad. Sci.* **23**, 331–332.

Bergelson, V.

[1986] A density statement generalizing Schur's theorem, *J. Combin. Theory A* **43**, 338–343.

Bergelson, V., and N. Hindman

[1988] A combinatorially Large Cell of a partition of *N*, *J. Combin. Theory A* **48**, 39–52.

Bergelson, V., and A. Leibman

[1995] Polynomial extensions of van der Waerden's and Szemerédi's theorems, *J. Amer. Math. Soc.* **8**, to appear.

Bergelson, V., W. Deuber and N. Hindman

[1991] Rado's theorem for finite fields, *Colloq. Math. Soc. János Bolyai* **60**.

Bergelson, V., W. Deuber, N. Hindman and H. Lefmann

[1995] Radó theorem for commutative rings, *J. Combin. Theory A*, to appear.

Berlekamp, E.R.

[1968] A construction for partitions which avoid long arithmetic progressions, *Canad. Math. Bull.* **11**, 409–414.

Bollobás, B.

[1978] *Extremal Graph Theory* (Academic Press, New York).

[1979] *Graph Theory* (Springer, Berlin).

[1985] *Random Graphs* (Academic Press, New York).

Bourgain, J.

[1986] A Szemerédi type theorem for sets of positive density in $\mathbb{R}^k$, *Israel J. Math.* **54**, 307–316.

Brauer, A.

[1928] Über Sequenzen von Potenzresten, *Sitzungsber. Preuss. Akad. Wiss.*, p. 9–16.

Brightwell, G.R., and Y. Kohayakawa

[1994] Ramsey properties of orientations of graphs, *Random Structures and Algorithms* **4**, 413–428.

Brown, T.C., and J.P. Buhler

[1982] A density version of a geometric Ramsey theorem, *J. Combin. Theory A* **32**, 20–34.

[1984] Lines imply spaces in density Ramsey Theory, *J. Combin. Theory A* **36**, 214–220.

Brown, T.C., F.R.K. Chung, P. Erdős and R.L. Graham

[1985] Quantitative forms of a theorem of Hilbert, *J. Combin. Theory A* **38**, 210–216.

Bukor, J.

[1994] A note on Folkman number $F(3,3;5)$, *Math. Slovaka* **44**(4), 479–480.

Burr, S.A., and P. Erdős

[1975] On the magnitude of generalised Ramsey number for graphs, *Colloq. Math. Soc. János Bolyai* **10**, 214–240.

[1976] Extremal Ramsey theory for graphs, *Utilitas Math.* **9**, 247–258.

Burr, S.A., P. Erdős and J.H. Spencer

[1975] Ramsey theorems for multiple copies of graphs, *Trans. Amer. Math. Soc.* **209**, 87–99.

Carlson, T.J.

[1988] Some underlying principles in Ramsey theory, *Discrete Math.* **68**, 117–169.

Carlson, T.J., and S.G. Simpson

[1984] A dual form of Ramsey's theorem, *Adv. in Math.* **53**, 265–290.

[1990] Topological Ramsey theory, in: *Mathematics of Ramsey Theory*, eds. J. Nešetřil and V. Rödl (Springer, Berlin) pp. 172–183.

Chen, G., and R.H. Schelp

[1993] Graphs with linearly bounded Ramsey numbers, *J. Combin. Theory B* **57**, 138–149.

Chintschin, A.J.
[1951] *Drei Perlen der Zahlentheorie* (Akademie Verlag, Berlin). Reprinted: 1984 (Verlag Harri Deutsch, Frankfurt).

Chung, F.R.K., and R.L. Graham
[1991] Quasi-random set systems, *J. Amer. Math. Soc.* **4**, 151–196.

Chung, F.R.K., and C.M. Grinstead
[1983] A survey of bounds for classical Ramsey numbers, *J. Graph Theory* **8**, 25–37.

Chung, F.R.K., R.L. Graham and R.M. Wilson
[1989] Quasirandom graphs, *Combinatorica* **9**, 345–362.

Chung, F.R.K., R. Cleve and P. Dagum
[1993] A note on constructive lower bounds for the Ramsey numbers $R(3,t)$, *J. Combin. Theory* **57**, 150–155.

Chvátal, V.
[1977] Three-complete graph Ramsey numbers, *J. Graph Theory* **1**, 93.

Chvátal, V., and F. Harary
[1972] Generalized Ramsey theory for graphs III: Small off-diagonal numbers, *Pacific J. Math.* **41**, 335–345.

Chvátal, V., V. Rödl, E. Szemerédi and W.T. Trotter
[1983] The Ramsey number of graph with bounded maximum degree, *J. Combin. Theory B* **34**, 239–243.

Coppersmith, D., and S. Winograd
[1990] Matrix multiplication via arithmetic progressions, *J. Symbolic Comput.* **9**, 251–280.

Deuber, W.
[1973] Partitionen und lineare Gleichungssysteme, *Math. Z.* **133**, 109–123.
[1975a] Partition Theorem für Graphen, *Math. Helv.* **50**, 311–320.
[1975b] Partition theorems for Abelian groups, *J. Combin. Theory A* **19**, 95–108.
[1982] On van der Waerden's theorem on arithmetic progressions, *J. Combin. Theory A* **32**, 115–118.

Deuber, W., and B.L. Rothschild
[1976] Categories without the Ramsey property, in: *Combinatorics*, eds. A. Hajnal and V.T. Sós, *Colloq. Math. Soc. János Bolyai* **18**, 225–249.

Deuber, W., and B. Voigt
[1982] Partitionseigenschaften endlicher affiner und projektiver Räume, *European J. Combin.* **3**, 329–340.
[1983] Der Satz von van der Waerden über Arithmetische Progressionen, *Jber. Deutsch. Math.-Verein* **85**, 66–85.

Deuber, W., B.L. Rothschild and B. Voigt
[1982] Induced partition theorems, *J. Combin. Theory A* **32**, 225–240.

Deuber, W., N. Hindman, I. Leader and H. Lefmann
[1995] Infinite partition regular matrices, *Combinatorica*, to appear.

Duffus, D., H. Lefmann and V. Rödl
[1995] Shift graph and lower bounds on Ramsey numbers $r_k(l;r)$, *Discrete Math.*, to appear.

Eaton, N., and V. Rödl
[1992] A canonical Ramsey theorem, *Random Structures and Algorithms* 3(4), 427–444.

Erdős, P.
[1947] Some remarks on the theory of graphs, *Bull. Amer. Math. Soc.* **53**, 292–294.
[1959] Graph theory and probability, *Canad. J. Math.* **11**, 34–38.
[1962] On the number of complete subgraphs contained in certain graphs, *Publ. Math. Inst. Hungar. Acad. Sci. VII, Ser. A* **3**, 459–464.
[1975] Problems and results on finite and infinite graphs, in: *Recent Advances in Graph Theory*, ed. M. Fiedler (Academia, Prague) pp. 183–192.
[1987] My joint work with R. Rado, in: *Surveys in Combinatorics 1987*, ed. C. Whitehead (Cambridge University Press, Cambridge) pp. 53–80.

Erdős, P., and R. Faudree
[1992] Size Ramsey functions, *Colloq. Math. Soc. János Bolyai* **60**, 219–238.

Erdős, P., and R. Graham

[1980] *Old and New Problems and Results in Combinatorial Number Theory, L'Enseignement Math.* **28**, 128pp.

Erdős, P., and A. Hajnal

[1966] On chromatic number of set systems, *Acta Math. Acad. Sci. Hungar.* **17**, 61 99.

[1977] On spanned subgraphs of graphs, in: *Beiträge zur Graphentheorie und deren Anwendungen* (Teubner, Leipzig) pp. 80–96.

Erdős, P., and L. Lovász

[1975] Problems and results on 3-chromatic hypergraphs and some related questions, in: *Infinite and Finite Sets*, eds. A. Hajnal, R. Rado and V.T. Sós (North-Holland, Amsterdam) pp. 609 628.

Erdős, P., and G. Mills

[1981] Some bounds for the Ramsey–Paris–Harrington numbers, *J. Combin. Theory A* **30**, 53–70.

Erdős, P., and R. Rado

[1950] A combinatorial theorem, *J. London Math. Soc.* **25**, 249–255.

[1952] Combinatorial theorems on classification of subsets of a given set, *Proc. London Math. Soc.* **2**, 417 439.

[1953] A problem on ordered sets, *J. London Math. Soc.* **28**, 426–438.

Erdős, P., and V.T. Sós

[1982] On Ramsey–Turán type theorems for hypergraphs, *Combinatorica* **2**, 289–295.

Erdős, P., and D. Szekeres

[1935] A combinatorial problem in geometry *Compos. Math.* **2**, 463–470.

Erdős, P., and P. Turán

[1936] On some sequences of integers, *J. London Math. Soc.* **11**, 261–264.

Erdős, P., A. Hajnal and R. Rado

[1965] Partition relations for cardinal numbers, *Acta Math. Acad. Sci. Hungar.* **16**, 93–196.

Erdős, P., R.L. Graham, P. Montgomery, B.L. Rothschild, J. Spencer and E.G. Straus

[1975a] Euclidean Ramsey theorems II, III, in: *Infinite and Finite Sets*, eds. A. Hajnal, R. Rado and V.T. Sós (North-Holland, Amsterdam) pp. 529–558, 559–584.

Erdős, P., A. Hajnal and L. Pósa

[1975b] Strong embeddings of graphs into colored graphs, in: *Infinite and Finite Sets*, eds. A. Hajnal, R. Rado and V.T. Sós, *Colloq. Math. Soc. János Bolyai* **10**, 585 595.

Erdős, P., J. Nešetřil and V. Rödl

[1983] On some problems related to partitions of edges of a graph, in: *Graphs and other Topics*, ed. M. Fiedler (Teubner, Leipzig) pp. 54–63.

Erdős, P., A. Hajnal, A. Máté and R. Rado

[1984a] *Combinatorial Set Theory: Partition Relations for Cardinals, Studies in Logic and the Foundations of Mathematics* (Akadémiai Kiadó/North-Holland, Budapest/Amsterdam).

Erdős, P., J. Nešetřil and V. Rödl

[1984b] Selectivity of hypergraphs, *Colloq. Math. Soc. János Bolyai* 37, 265–284.

Erdős, P., R. Faudree, C.C. Rousseau and R.H. Schelp

[1985] Multipartite graph–sparse graph Ramsey numbers, *Combinatorica* 5, 311–318.

Erdős, P., P. Frankl and V. Rödl

[1986] The asymptotic number of graphs not containing a fixed subraph and a problem for hypergraphs having no exponent, *Graphs and Combin.* **2**, 113–121.

Erdős, P., R. Faudree, C.C. Rousseau and R.H. Schelp

[1989] Multipartite graph–tree Ramsey numbers, *Ann. New York Acad. Sci.* **576**, 146–154.

Erdős, P., A. Hajnal, M. Simonovits, V.T. Sós and E. Szemerédi

[1993] Turán–Ramsey theorems and simple asymptotically extremal structures, *Combinatorica* **13**(1), 31–56.

Erdős, P., M. Loebl and V.T. Sós

[1995] Discrepancy of trees, *Studia Sci. Math. Hungar.*, to appear.

Erickson, M.

[1993] An upper bound for the Folkman Number $F(3,3;5)$, *J. Graph Theory* **17**(6), 679–681.

Fagin, R.

[1976] Probabilities finite models, *J. Symbolic Logic* **41**, 50–58.

Faudree, F., and M. Simonovits

[1992] Ramsey problems and their connection to Turán type extremal problems, *J. Graph Theory* **16**, 2550.

Folkman, J.

[1970] Graph with monochromatic complete subgraphs in every edge coloring, *SIAM J. Appl. Math.* **18**, 19–24.

Frankl, P.

[1977] A constructive lower bound for Ramsey numbers, *Ars Combin.* **2**, 297–302.

[1990] Constructive Ramsey bounds and intersection theorems for sets, in: *Mathematics of Ramsey Theory*, eds. J. Nešetřil and V. Rödl (Springer, Berlin) pp. 53–56.

Frankl, P., and V. Rödl

[1986] All triangles are Ramsey, *Trans. Amer Math. Soc.* **297**, 777–779.

[1990] A partition property of simplices in Euclidean space, *J. Amer. Math. Soc.* **3**(1), 1–7.

Frankl, P., and R.M. Wilson

[1981] Intersection theorems with geometric consequences, *Combinatorica* **1**, 357–368.

Frankl, P., R.L. Graham and V. Rödl

[1987] Induced restricted Ramsey theorems for spaces, *J. Combin. Theory A* 120–128.

[1988] Quantitative theorems for regular systems of equations, *J. Combin. Theory A* **47**, 246–261.

[1989] On the distribution of monochromatic configurations, in: *Partitions and Irregularities*, eds. G. Halász and V.T. Sós (Springer, Berlin) pp. 71–87.

Fürstenberg, H.

[1977] Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions, *J. Anal. Math.* **31**, 204–256.

[1981] *Recurrence in Ergodic Theory and Combinatorial Number Theory* (Princeton University Press, Princeton, NJ).

Fürstenberg, H., and Y. Katznelson

[1978] An ergodic theorem for commuting transformations, *J. Analyse Math.* **31**, 204–256.

[1985] An ergodic Szemerédi theorem for *IP*-systems and combinatorial theory, *J. Analyse Math.* **45**, 117–168.

[1989] Idempotents in compact semigroups and Ramsey theory, *Israel J. Math.* **68**, 257–270.

[1991] A density version of the Hales–Jewett theorem, *J. Analyse Math.* **57**, 64–119.

Fürstenberg,, H., and B. Weiss

[1978] Topological dynamics and combinatorial number theory, *J. Analyse Math.* **38**, 61–85.

Glebskii, Y.V., D.I. Kogan, M.I. Liogonkii and V.A. Talanov

[1969] Range and degree of realizability of formulas in the restricted predicate calculus, *Cybernetics* **5**, 142–154.

Goodman, A.W.

[1959] On sets of acquaintances and strangers at any party, *Amer. Math. Monthly* **66**, 778–783.

Graham, R.L.

[1968] On edgewise 2-colored graphs with monochromatic triangles and containing no complete hexagon, *J. Combin. Theory* 4, 300.

[1980] On partitions of $\mathbb{E}^n$, *J. Combin. Theory A* 28, 89–97.

[1981] Rudiments of Ramsey Theory, *CBMS Regional Conf. in Math.*, No. 45 (American Mathematical Society, Providence, RI).

[1990] Topics in Éuclidean Ramsey theory, in: *Mathematics of Ramsey Theory*, eds. J. Nešetřil and V. Rödl (Springer, Berlin) pp. 200–213.

[1995] Recent trends in Euclidean Ramsey theory, *Recent Trends in Discrete Mathematics, Discrete Math.*, to appear.

Graham, R.L., and V. Rödl

[1987] Numbers in Ramsey Theory, in: *Surveys in Combinatorics*, ed. C. Whitehead (Cambridge University Press, Cambridge).

Graham, R.L., and B.L. Rothschild

[1971] Ramsey's theorem for n-parameter sets, *Trans. Amer. Math. Soc.* **159**, 257–292.

[1974] A short proof of van der Waerden's theorem on arithmetic progressions, *Proc. Amer. Math. Soc.* **42**, 356–386.

Graham, R.L., K. Leeb and B.L. Rothschild

[1972] Ramsey's theorem for a class of categories, *Adv. in Math.* **8**, 417–433.

Graham, R.L., B.L. Rothschild and J.H. Spencer

[1980] *Ramsey Theory* (Wiley, New York). 2nd Edition: 1990.

Graver, J.E., and J. Yackel

[1968] Some graph theoretic results associated with Ramsey's theorem, *J. Combin. Theory* **4**, 125–175.

Greenwood, R.E., and A.M. Gleason

[1955] Combinatorial relations and chromatic graphs, *Canad. J. Math.* **7**, 1–7.

Grinčuk, M.I.

[1989] On the complexity of realizations of symmetric Boolean functions by contact schemas, *Dokl. Akad. Nauk* **309**(4), 787–791.

Grinstead, C., and S. Roberts

[1982] On the Ramsey Numbers $R(3,8)$ and $R(3,9)$, *J. Combin. Theory B* **33**, 27–51.

Hales, A.W., and R.I. Jewett

[1963] Regularity and positional games, *Trans. Amer. Math. Soc.* **106**, 222–229.

Harborth, H.

[1978] Konvexe Fünfecke in ebenen Punktmengen, *Elemente Math.* **335**, 116–118.

Hilbert, D.

[1892] Über die Irreducibilität Ganzer Rationaler Functionen mit Ganzzahligen Coefficienten, *J. Reine Angew. Math.* **110**, 104–129.

Hindman, N.

[1974] Finite sums from sequences within cells of a partition of N, *J. Combin. Theory A* **17**, 1–11.

[1979] Ultrafilters and combinatorial number theory, in: *Number Theory, Carbondale*, ed. M. Nathanson, *Lecture Notes in Mathematics*, Vol. 751, pp. 119–184.

Horton, J.D.

[1983] Sets with no empty convex 7-gons, *Canad. Math. Bull* **26**(4), 482–484.

Irving, R.

[1973] On a bound of Graham and Spencer for a graph-coloring constant, *J. Combin. Theory B* **15**, 200–203.

Ježek, J., and J. Nešetřil

[1983] Ramsey varieties, *European J. Combin.* **4**, 132–147.

Jockush, C.G.

[1972] Ramsey's theorem and recursion theory, *J. Symbolic Logic* **37**, 268–280.

Ketonen, J., and R. Solovay

[1981] Rapidly growing Ramsey functions, *Ann. of Math.* **113**, 267–314.

Kim, J.H.

[1995] The Ramsey Number $R(3,t)$ has order of magnitude $t^2/\log t$, *Random Structures and Algorithms*, to appear.

Kirby, L., and J. Paris

[1982] Accessible independence results for Peano Arithmetic, *Bull. London Math. Soc.* **14**.

Klee, P.

[1923] Doppelzelt, *Coll. Rosengart, Lucerne.*

Kolaitis, P.G., and M. Vardi

[1987] The decision problem for the probabilities of higher-order properties, in: *19th Annu. ACM Symp. (STOC)*, pp. 425–435.

Komlós, J., J. Pintz and E. Szemerédi

[1981] On Heilbronns triangle problem, *J. London Math. Soc.* **24**, 385–396.

Kříž, I.

[1987] Large independent sets in shift-invariant graphs, *Graphs and Combin.* **3**, 145–158.

[1991] Permutation groups in Euclidean Ramsey theory, *Proc. Amer. Math. Soc.* **112**, 899–907.

[1992a] All trapezoids are Ramsey, *Discrete Math.* **108**, 59–62.

[1992b] Equivariant cohomology and lower bounds for chromatic number, *Trans. Amer. Math. Soc.* **333**(2), 567–577.

Kříž, I., and R. Thomas

[1990] Ordinal types in Ramsey theory and well-partial-ordering theory, in: *Mathematics of Ramsey Theory*, eds. J. Nešetřil and V. Rödl (Springer, Berlin) pp. 57–95.

Kruskal, J.B.

[1960] Well-quasi-ordering, the tree theorem, and Vázsonyi's conjecture, *Trans. Amer. Math. Soc.* **95**, 210–225.

Lachlan, A.H., and R.E. Woodrow

[1980] Countable ultrahomogeneous graphs, *Trans. Amer. Math. Soc.* **262**, 51–94.

Larmann, D., J. Matoušek, J. Pach and J. Töröcsik

[1992] A Ramsey-type result for planar convex sets, to appear.

Leeb, K.

[1973] *Vorlesungen über Pascaltheorie* (Erlangen, 1973).

Lefmann, H.

[1986] A canonical version for partition regular systems of linear equations, *J. Combin. Theory A* **41**, 95–104.

[1991] On partition regular systems of equations, *J. Combin. Theory A* **58**, 35–53.

Lefmann, H., and V. Rödl

[1993] On canonical Ramsey numbers for complete graphs, *J. Combin. Theory B* **58**, 1–13.

[1995] On Erdős–Rado numbers, *Combinatorica*, to appear.

Lewis, H.R.

[1979] *Unsolvable Classes of Quantificational Formulas* (Addison-Wesley, Reading, MA).

Loebl, M.

[1988] Hercules and Hydra, *Comment. Math. Univ. Carolin.* **29**(1), 85–95.

[1992] Unprovable combinatorial theorems, *Discrete Math.* **108**, 333–342.

Loebl, M., and J. Matoušek

[1987] On undecidability of the weakened Kruskal theorem, *Contemporary Math. AMS*, pp. 275–280.

Loebl, M., and J. Nešetřil

[1991] Unprovable combinatorial statements, in: *Surveys in Combinatorics*, Vol. 166 (London Mathematical Society, London) pp. 119–160.

[1992] An unprovable Ramsey type theorem, *Proc. Amer. Math. Soc.* **116**(3), 819–824.

Lovász, L.

[1968] On the chromatic number of finite set-systems, *Acta Math. Acad. Sci. Hungar.* **19**, 59–67.

Lubotzky, A., R. Phillips and P. Sarnak

[1988] Ramanujan graphs, *Combinatorica* **8**(3), 261–277.

Margulis, G.A.

[1975] Explicit constructions of concentrators, *Problemy Peredachi Informatsii* 9(4), 71–80 [*Problems Inform. Transmission*, pp. 325–332].

Matoušek, J., and V. Rödl

[1995] On Ramsey sets in spheres, *J. Combin. Theory A*, to appear.

McKay, B.D., and S.P. Radziszowski

[1991] The first classical Ramsey number for hypergraphs is computed, in: *Proc. 2nd Annu. ACM–SIAM Symp. on Discrete Algorithms, SODA '91*, pp. 304–308.

McKay, B.D., and Zhang Ke Min

[1992] The value of the Ramsey number $R(3,8)$, *J. Graph Theory* **16**, 183–202.

Mills, G.

[1985] Ramsey–Paris–Harrington numbers for graphs, *J. Combin. Theory A* **38**, 30–37.

Nash-Williams, C.St.J.A.

[1965] On well-quasi ordering transfinite sequences, *Proc. Cambridge Philos. Soc.* **61**, 33–39.

Nešetřil, J.

[1984] Some non-standard Ramsey-like applications, *Theoret. Comput. Sci.* **84**, 3–15.

[1989] There are 4 types of Ramsey classes of graphs, *J. Combin. Theory B* **46**(2), 127–132.

[1994] On ordered graphs and graph orderings, *Discrete Appl. Math.* **51**, 113–116.

Nešetřil, J., and V. Rödl

[1966] Partitions of vertices, *Comment. Math. Univ. Carolin.* **17**, 675–681.

[1975a] Type theory of partition properties of graphs, in: *Recent Advances in Graph Theory*, ed. M. Fiedler (Academia, Prague) pp. 405–412.

[1975b] Partitions of subgraphs, in: *Recent Advances in Graph Theory*, ed. M. Fiedler (Academia, Prague) pp. 413–423.

[1976] Van der Waerden theorem for sequences of integers not containing an arithmetic progression of k-terms, *Comment. Math. Univ. Carol.* **17**(4), 675–682.

[1977a] A structural generalization of the Ramsey Theorem, *Bull. Amer. Math. Soc.* **83**(1), 127–128.

[1977b] Partition of relational and set-systems, *J. Combin. Theory A* **22**, 289–312.

[1978a] Partition (Ramsey) theory – a survey, *Colloq. Math. Soc. János Bolyai* **18**, 759–792.

[1978b] A probabilistic graph theoretical method, *Proc. Amer. Math. Soc.* **72**, 417–421.

[1979a] A short proof of the existence of highly chromatic graphs without short cycles, *J. Combin. Theory B* **27**, 225–227.

[1979b] Partition theory and its applications, in: *Surveys in Combinatorics* (Cambridge University Press, Cambridge) 96–156.

[1980] Dual Ramsey-type theorems, in: *Proc. VIIIth Winterschool on Abstract Analysis, Prague 1980*, ed. Z. Frolík, pp. 121–123.

[1981] A short proof of the existence of restricted Ramsey graphs by means of a partite construction, *Combinatorica* **1**(2), 199–202.

[1982] Two proofs of the Partition property of set systems, *European J. Combin.* **3**, 347–352.

[1983a] Ramsey classes of set systems, *J. Combin. Theory A* **34**, 183–201.

[1983b] Another proof of Folkman–Rado–Sanders theorem, *J. Combin. Theory A* **34**(1), 108–109.

[1984] Sparse Ramsey graphs, *Combinatorica* **4**(1), 71–78.

[1985] Two remarks on Ramsey theorem, *Discrete Math.* **53**, 339–341.

[1987a] Strong Ramsey theorems for Steiner systems, *Trans. Amer. Math. Soc.* **303**, 183–192.

[1987b] Partite construction and Ramseyian theorems for sets, numbers and spaces, *Comment. Math. Univ. Carol.* **28**, 569–580.

[1989] The partite construction and Ramsey set systems, *Discrete Math.* **75**, 327–334.

[1990a] *Mathematics of Ramsey Theory* (Springer, Berlin).

[1990b] The partite construction and Ramsey space systems, in: *Mathematics of Ramsey Theory*, eds. J. Nešetřil and V. Rödl (Springer, Berlin) pp. 98–112.

Nešetřil, J., and R. Thomas

[1987] WQO, long games and a combinatorial study of unprovability, in: *Logic and Combinatorics, Contemporary Math., Amer. Math. Soc., Providence*, pp. 281–293.

Nešetřil, J., and P. Valtr

[1994] A Ramsey type result in the plane, *Combinatorics, Probability and Computing* **3**, 127–135.

Nešetřil, J., H.J. Prömel, V. Rödl and B. Voigt

[1985] Canonizing ordering theorems for Hales Jewett structures, *J. Combin. Theory A* **40**, 394–408.

Nilli, A.

[1990] Shelah's proof of the Hales–Jewett Theorem, in: *Mathematics of Ramsey Theory*, eds. J. Nešetřil and V. Rödl (Springer, Berlin) pp. 150–151.

Paris, J.

[1990] Combinatorial unprovable theorems, in: *Mathematics of Ramsey Theory*, eds. J. Nešetřil and V. Rödl (Springer, Berlin) pp. 232–245.

Paris, J., and L. Harrington

[1977] A mathematical incompleteness in Peano Arithmetic, in: *Handbook of Mathematical Logic*, ed. J. Bairwise (North-Holland, Amsterdam) pp. 1133–1142.

Pelant, J., and V. Rödl

[1992] On coverings of infinite dimensional metric spaces, in: *Frolik's Memorial Volume, Topics in Discrete Mathematics*, Vol. 8, ed. J. Nešetřil (North-Holland, Amsterdam) pp. 75–81.

Pomerance, C.

[1980] Collinear subsets of lattice point sequence – an analog of Szemerédi's theorem, *J. Combin. Theory A* **28**(2), 140–149.

Pósa, L.

[1976] Hamiltonian circuits in random graphs, *Discrete Math.* **14**, 359–364.

Prömel, H.J.

[1985] Induced partition properties of combinatorial cubes, *J. Combin. Theory A* **39**, 177–208.

Prömel, H.J., and B. Voigt

[1981] Recent results in partition (Ramsey) theory for finite lattice, *Discrete Math.* **35**, 185–198.

[1983] Canonical parition theorems for parameter sets, *J. Combin. Theory A* **35**, 309–327.

[1985] *Canonizing Ramsey Theory*, Preprint (Universität Bonn).

[1988] A sparse Graham–Rothschild theorem, *Trans. Amer. Math. Soc.* **309**(1), 113–138.

[1990] Graham–Rothschild parameter sets, in: *Mathematics of Ramsey Theory*, eds. J. Nešetřil and V. Rödl (Springer, Berlin) pp. 113–149.

Pudlák, P.

[1984] Bounds for Hodes Specker theorem, in: *Logic and Machines: Decision Problems and Complexity, Lecture Notes in Computer Science*, Vol. 171 (Springer, Berlin) pp. 421–445.

Rado, R.

[1933] Studien zur Kombinatorik, *Math. Z* **36**, 425–480.

[1986] Note on canonical partitions, *Bull. London Math. Soc.* **18**, 123–126. Reprinted: 1990, in: *Mathematics of Ramsey Theory*, eds. J. Nešetřil and V. Rödl (Springer, Berlin) pp. 29–32.

Radziszowski, S.P.

[1993] *Small Ramsey Numbers*, Preprint (Rochester Institute of Technology, February).

Ramsey, F.P.

[1930] On a problem for formal logic, *Proc. London Math. Soc.* **30**, 264–286.

[1978] in: *Foundations*, ed. D.H. Mellor (Routledge and Kegan Paul, London).

Ray-Chaudhuri, D.K., and R.M. Wilson

[1975] On t-designs, *Osaka J. Math.* **12**, 735–744.

Rödl, V.

[1976] A generalization of Ramsey theorem, in: *Graphs, Hypergraphs and Block Systems, Zielona Gora*, pp. 211–220.

[1982] Note on finite Boolean algebras, *Acta Polytechnica, CVUT (Praha)* pp. 47–49.

[1986a] *Upper bounds on Ramsey numbers $R(k,l)$*, Unpublished Manuscript.

[1986b] On universality of graphs with uniformly distributed edges, *Discrete Math.* **59**, 125–134.

[1990] On Ramsey families of sets, *Graphs and Combin.* **6**, 187–195.

[1991] Recent results in Ramsey Theory, in: *Proc. Int. Congr. of Mathematicians, 1990* (Springer, Berlin) pp. 1455–1466.

Rödl, V., and A. Ruciński

[1993] Lower bounds on probability thresholds for Ramsey properties, in: *Combinatorics*, eds. D. Miklós, V.T. Sós and T. Szonyi (J. Bolyai Math. Soc., Budapest).

[1995] Threshold functions for Ramsey properties, *J. Amer. Math. Soc.* **8**, to appear.

Rödl, V., and N. Sauer

[1992] The Ramsey property of graphs which exclude a given graph, *Canad. J. Math.* **44**(5), 1050–1060.

Rödl, V., and R. Thomas

[1995] Arrangeability and clique subdivisions, in: *Mathematics of Paul Erdős*, eds. R.L. Graham and J. Nešetřil, to appear.

., V., and P. Winkler
)89] Ramsey type theorem for ordering of a graph, *SIAM J. Discrete Math.* 2(3), 402–406.

Rödl, V., N. Sauer and X. Zhu
[1995] Ramsey families which exclude a graph, to appear.

Roth, K.
[1953] On certain sets of integers, *J. London Math. Soc.* **28**, 104–109.

Ruzsa, I.Z., and E. Szemerédi
[1978] Triple systems with no six points carrying three triangles, *Colloq. Math. Soc. János Bolyai* **18**, 939–945.

Salem, R., and D.C. Spencer
[1942] On sets of integers which contain no three terms in arithmetical progression, *Proc. Nat. Acad. Sci. USA* **28**, 561–563.

Sanders, J.H.
[1968] *A generalization of Schur's theorem*, Doctoral Dissertation (Yale University).

Schur, I.
[1916] Über die Kongruenz $x^m + y^m \equiv z^m$ (mod$p$), *Jber. Deutsch. Math.-Verein* **25**, 114–117.
[1973] *Gesammelte Abhandlungen*, eds. A. Brauer and H. Rohrbach (Springer, Berlin).

Shearer, J.B.
[1983] A note on the independence number of a triangle free-graph, *Discrete Math.* **46**, 83–87.

Shelah, S.
[1988] Primitive recursive bounds for van der Waerden numbers, *J. Amer. Math. Soc.* **1**, 683–697.
[1994] *Finite Canonization*, Preprint.

Sidorenko, A.F.
[1991] An upper bound on the Ramsey number $R(K_3, G)$ depending only on the size of the graph G, *J. Graph Theory* **15**, 15–17.
[1994] The Ramsey Number of an N-edge graph versus triangle is at most $2N + 1$, *J. Combin. Theory B*, to appear.

Simonovits, M., and V.T. Sós
[1991] Szemerédi's partition and quasirandomness, *Random Structures and Algorithms* **2**, 1–10.

Simpson, S.G.
[1985] Non-provability of certain combinatorial properties of finite trees, in: *Harvey Friedman's Research of the Foundation of Mathematics*, eds. L. Harrington, M. Morley, A. Scedror and S.G. Simpson (North-Holland, Amsterdam) pp. 87–117.

Sparks, N.
[1993] *Definable Cases of the Hales Jewett Theorem*, Preprint.

Spencer, J.
[1975a] Ramsey's theorem – a lower bound, *J. Combin. Theory A* **18**, 108–115.
[1975b] Restricted Ramsey configurations, *J. Combin Theory A* **19**, 278–286.
[1977] Asymptotic lower bounds for Ramsey functions, *Discrete Math.* **20**, 69–76.
[1979] Ramsey's theorem for spaces, *Trans. Amer. Math. Soc.* **249**, 363–371.
[1981] Extremal problems, partition theorems, symmetric hypergraphs, *Combinatorica* **1**, 303–337.
[1983] Ramsey theory and Ramsey theoreticians, *J. Graph Theory* 7, 15–23.
[1985] Four roads to the Ramsey function, *Ann. Discrete Math.*, pp. 243–250.
[1988] Three hundred million points suffice, *J. Combin. Theory A* **49**, 210–217.

Szemerédi, E.
[1975] On sets of integers containing no k elements in arithmetic progression, *Acta Arithm.* **27**, 199–245.
[1976] Regular partitions of graphs, in: *Proc. Colloq. Int. CNRS* (CNRS, Paris) pp. 399–401.

Taylor, A.D.
[1981] Bounds for the disjoint union theorem, *J. Combin. Theory A* **30**, 339–344.
[1982] A note on van der Waerden theorem, *J. Combin. Theory A* **33**, 215–219.

Thomasson, A.
[1987] Random graphs, strongly regular graphs and pseudorandom graphs, in: *Survey in Combinatorics 1987*, ed. C. Whitehead (Cambridge University Press, Cambridge) pp. 173–196.

[1988] An upper bound for some Ramsey numbers, *J. Graph Theory* **12**, 509–517.

Trakhtenbrot, B.A.

[1950] The impossibility of an algorithm for the decision problem for finite models, *Dokl. Akad. Nauk SSR* **70**, 569–572.

Tutte, W.T. (Blanche Descartes)

[1954] Solution to Advanced Problem Nr. 4526, *Amer. Math. Monthly* **61**, 352.

Valtr, P.

[1992] Convex independent sets and 7-holes in restricted planar point sets, *Discrete Comput. Geom.* **7**, 135–152.

van der Waerden, B.L.

[1927] Beweis einer Baudetschen Vermutung, *Nieuw Arch. Wisk.* **15**, 212–216.

[*1971*] *How the proof of Baudet's Conjecture was found, in: Studies in Pure Mathematics,* ed. L. Mirsky (Academic Press, New York) pp. 251–260.

Vilfan, B.

[1976] Lower bounds for the size of expressions for certain functions in d-ary logic, *Theor. Comput. Sci.* **2**, 249–269.

Voigt, B.

[1980] The partition problem for finite abelian groups, *J. Combin. Theory A* **28**, 257–271.

Wainer, S.

[1970] A classification of the ordinal recursive functions, *Arch. Math. Logic* **13**.

Walker, K.

[1971] An upper bound for the Ramsey Number $M(5,4)$, *J. Combin. Theory A*, , pp. 1–10.

Zykov, A.

[1949] On some properties of linear complexes, *Mat. Sbornik N.S.* **24**, 163–188 [1952, *AMS Transl.* **79**]. Reprinted: *1962, Translations Series 1, Vol. 7* (AMS, Providence, RI) *pp.* 418–449.

CHAPTER 26

Discrepancy Theory

József BECK

Rutgers University, Department of Mathematics, New Brunswick, NJ 08903, USA

Vera T. SÓS

Mathematical Research Institute of the Hungarian Academy of Sciences, Reáltanoda utca 13–15,
Budapest 1364, Hungary

Contents

HANDBOOK OF COMBINATORICS
Edited by R. Graham, M. Grötschel and L. Lovász

1. Introduction

The concept of *uniformly distributed sequences* and *sets* plays a fundamental role in many branches of mathematics (measure theory, ergodic theory, diophantine approximation, mathematical statistics, discrete geometry, numerical integration, etc.) This chapter explores the combinatorial background of many of these results. See also the survey article of Sós (1983b), and the monograph by Beck and Chen (1987).

Measure theoretic discrepancy results are accumulated in two complementary chapters of number theory, called *uniform distribution* and *irregularities of distribution*. The object of these theories is to measure the uniformity (or non-uniformity) of sequences and point distributions. For instance: how uniformly can N points in the unit cube be distributed relative to a given family of "nice" sets (e.g., boxes with sides parallel to the coordinate axes, rotated boxes, balls, all convex sets, etc.). The theory was initiated by the following theorem of Aardenne-Ehrenfest (Van der Corput's conjecture): for every infinite sequence of reals in $[0, 1]$ and for every $k > 0$, there exists a beginning section $(x_1, \ldots, x_n)$ of the sequence and a subinterval (α, β) such that the number of elements of this beginning section in this subinterval differs from $n(\beta - \alpha)$ (the number one expects) by at least k. The best possible effective result on this problem is due to Schmidt; it is equivalent to the following basic result in the theory of uniform distribution.

Theorem 1.1 (Schmidt 1972). *Let P be an arbitrary set of N points in the unit square $[0, 1)^2$. Then there exists a rectangle $B \subset [0, 1)^2$ with sides parallel to the coordinate axes such that*

$$\Big| |P \cap B| - N \operatorname{area}(B) \Big| > c \log N$$

(*where c is an absolute constant*).

The left-hand side of this inequality measures the "discrepancy" (deviation from the uniform distribution) of P in B. As a fascinating fact, we mention that balls have much greater discrepancy than boxes with sides parallel to the axes. Now we have a good understanding of this phenomenon, as we shall see later.

The object of combinatorial discrepancy theory is to color a set with two or more colors so that each set in a given family be colored as uniformly as possible. As a beautiful example, we mention Roth's theorem on long arithmetic progressions.

Theorem 1.2 (Roth 1964). *For any partition of the integers $1, 2, \ldots, N$ into two sets S_1 and S_2, there exists an arithmetic progression $P = \{a, a + d, \ldots, a + kd\} \subset \{1, 2, \ldots, N\}$ such that*

$$\Big| |P \cap S_1| - |P \cap S_2| \Big| > \tfrac{1}{20} N^{1/4} .$$

It took more than a decade to realize the close relationship between these areas. We can now say that they represent the continuous and the discrete aspects of the very same coherent theory. A general form of these problems is the following: given a measure space, approximate the measure on a subfamily of the measurable sets by a measure where each point has measure 0 or 1. Nontrivial "transference theorems" help to transform combinatorial and measure theoretic results into each other.

Compare Roth's theorem also to the following fundamental results of Ramsey theory (see chapter 25).

Theorem 1.3 (Van der Waerden 1927). *For any integers k and r there exists an $W(k,r)$ such that if $N > W(k,r)$ then for every r-coloring of $\{1,2,\dots,N\}$ there exists a monochromatic arithmetic progression of length k.*

Theorem 1.4 (Ramsey 1930). *For any integers t and r there exists an $R(t,r)$ such that if $n > R(t,r)$ and the edges of K_n are r-colored, then there must be a monochromatic K_t.*

These theorems have the same structure as Roth's: given an underlying set S and a family of subsets of this set, the claim is that the underlying set has no partition which splits each set contained in the given family "reasonably well" (only in this case any proper splitting is accepted).

Discarding the special structure of the system we can formulate the basic problem in combinatorial discrepancy theory. Let $S = \{x_1, \dots, x_n\}$ be a finite set and $\mathcal{H} = \{A_1, \dots, A_m\}$, a family of subsets of S. Our goal is to find a partition $S = S_1 \cap S_2$, $S_1 \cap S_2 = \emptyset$ that splits each set in the family $\mathcal{H}$ as equally as possible. In other words, we want to find the least integer D for which there exists a 2-coloring of the underlying set such that in each A_i, the difference between the numbers of red and blue elements is at most D.

Often we shall describe the partition by a function $f: S \to \{-1, 1\}$. Then the *discrepancy* of $\mathcal{H}$ is defined by

$$\mathcal{D}(\mathcal{H}) = \min_f \max_{1 \le j \le m} \left| \sum_{x_i \in A_j} f(x_i) \right| ,$$

where the minimum is taken over all functions $f: S \to \{-1, 1\}$.

Best and worst families. Although the systematic investigation of combinatorial discrepancy started just a few years ago, there is a fundamental old result which characterizes the "best" families, those for which $\mathcal{D}(\mathcal{H}) \le 1$, and this is inherited to subhypergraphs. These are the unimodular hypergraphs, whose theory was developed for its importance in integer programming (see chapter 30).

A hypergraph $\mathcal{H}$ is *unimodular*, if its incidence matrix A is totally unimodular (i.e., every square submatrix of A has determinant 0, +1 or −1). See chapters 7 and 30 for examples of such hypergraphs; here we mention hypergraphs whose edges are the vertex sets of directed paths in an arborescence. For $X \subseteq S$, the *restriction* $\mathcal{H}_X$ is defined as the family $\{A \cap X \mid A \in \mathcal{H}\}$.

Theorem 1.5 (Ghouila-Houri 1962). *$\mathcal{H}$ is unimodular iff $\mathcal{D}(\mathcal{H}_X) \leq 1$ for all restrictions $\mathcal{H}_X$ of $\mathcal{H}$.*

Unimodular hypergraphs have the following stronger property.

Theorem 1.6. *If $\mathcal{H} = (V, E)$ is unimodular then for any $p \in [-1, 1]^V$ there exist $\varepsilon \in \{-1, 1\}^V$ such that for every $A \in E$,*

$$\left| \sum_{i \in A} (\varepsilon_i - p_i) \right| \leq 1 .$$

Informally, an arbitrary weight distribution on S can be very well approximated with 0–1 weights.

Furthermore, we have the following.

Theorem 1.7. *If $\mathcal{H}$ is unimodular, then for every $r > 1$ there exists an r-equipartition $S = S_1 \cup \cdots \cup S_r$ so that for every $A \in \mathcal{H}$ and $1 \leq j \leq r$,*

$$\left\lfloor \frac{A}{r} \right\rfloor \leq |A \cap S_j| \leq \left\lceil \frac{A}{r} \right\rceil .$$

The "*worst*" families from the point of view of discrepancy are the "non-2-colorable families", i.e., families with chromatic number $\chi(\mathcal{H}) > 2$. (Recall from chapter 7 that a hypergraph is non-2-colorable iff for any partition $S = S_1 \cup S_2$ there exists an $A \in \mathcal{H}$ so that $A \subseteq S_1$ or $A \subseteq S_2$.) An r-uniform hypergraph is 2-colorable if and only if its discrepancy is less than r. (Note that this remark also shows that the computation of the discrepancy of a hypergraph is NP-hard.)

One of the most extensively studied field of combinatorics is Ramsey theory, which can be viewed as the theory of non-2-colorable families (see chapter 7). Many of the results and problems there are relevant to our subject.

Considering the results in Ramsey theory we must realize the white spots and gaps in discrepancy theory. A large variety of Ramsey-type results are available not only for graphs and hypergraphs but for different structures like vector spaces, combinatorial lines, parameter-sets, groups, euclidean spaces, topological spaces, sets of solutions of linear systems, etc. However, an analogous discrepancy theory is missing for most of these structures.

We can say that in the class of hypergraphs unimodular families are at one (at the "good") end and non-2-colorable families at the other ("bad") end. We conclude this section with an example of A.J. Hoffman showing that the union of two unimodular (so best!) families can be non-2-colorable (so worst!).

Example 1.8 (Hoffman 1987). Let T be an arbitrary arborescence rooted at r. Let $\mathcal{H}_1$ consist of the arc-sets of directed paths in T from r to a leaf. Let $\mathcal{H}_2$ consist of the sets $B(x)$, where $B(x)$ is the set of edges with their tails at node x, for each non-leaf node x. Obviously $\mathcal{H}_1$ and $\mathcal{H}_2$ are unimodular, but $\mathcal{H}_1 \cup \mathcal{H}_2$ is not even 2-colorable. (Note that we can choose the tree so that $\mathcal{H}_1 \cup \mathcal{H}_2$ is k-uniform for a given k.)

A very simple unimodular hypergraph is the hypergraph of all intervals in a permutation (a totally ordered set). How large can be the discrepancy of the union of such hypergraphs? For two permutations, the discrepancy is at most 2; but the following problem, due to Beck, has been open for quite a while.

Problem 1.9. Is it true that the hypergraph consisting of the intervals of three permutations of a set X has discrepancy $O(1)$, independent of $|X|$?

Recently Bohus (1990) gave the upper bound $O(\log |X|)$ for this discrepancy, not only for three, but for any constant number of permutations.

2. Bounds on $\mathcal{D}(\mathcal{H})$

Many of the results in this section have applications in different fields. In fact, many of the problems originated in different branches of mathematics.

There is a trivial upper bound on the combinatorial discrepancy:

$$\mathcal{D}(\mathcal{H}) \leq \max_{A \in \mathcal{H}} |A| .$$

If $\mathcal{H}$ is k-uniform (i.e., $|A| = k$ for all $A \in \mathcal{H}$) then equality holds iff $\mathcal{H}$ is a not 2-colorable.

To bound the discrepancy in terms of the number of edges $m = |\mathcal{H}|$, observe that a pair of vertices contained in the same set of edges can be deleted without decreasing the discrepancy. Repeating this we end up with a hypergraph in which every edge has at most $2^m - 1$ elements and hence

$$\mathcal{D}(\mathcal{H}) \leq 2^m - 1 .$$

This upper bound can be easily improved. The first result in this direction was the theorem of Olson and Spencer (1978) where they proved the upper bound

$$\mathcal{D}(\mathcal{H}) \leq cm^{1/2} \log m .$$

The best possible result is the following.

Theorem 2.1 (Spencer 1985). *For every $\mathcal{H}$ with $|\mathcal{H}| = m$*

$$\mathcal{D}(\mathcal{H}) \leq 6m^{1/2} .$$

For a proof, which is an involved application of the probabilistic method, see chapter 33. This result is best possible (up to a constant): if an Hadamard matrix of order $m+1$ exists, then there exists a hypergraph $\mathcal{H}$ with $|\mathcal{H}|=m$ such that $\mathcal{D}(\mathcal{H}) \geqslant \frac{1}{2}m^{1/2}$ (see Corollary 2.11).

Spencer's theorem has interesting applications in Fourier analysis to "Rudin–Shapiro sequences" (see Spencer 1985), and to Littlewood's problem on "flat polynomials" (see Beck 1991b).

It is somewhat surprising that there is an upper bound on $\mathcal{D}(\mathcal{H})$ depending only on the maximum degree $\Delta(\mathcal{H}) = \max_{x\in S} |\{A \in \mathcal{H} : x \in A\}|$.

Theorem 2.2 (Beck and Fiala 1981). *Let $\mathcal{H}$ be a finite hypergraph. Then*

$$\mathcal{D}(\mathcal{H}) < 2\Delta(\mathcal{H}) .$$

In fact, we have the following more general result.

Theorem 2.2′. *Let us associate with every $i \in S$ a real number $p_i \in [-1, +1]$. Then there exist $\varepsilon_i \in \{-1, +1\}$ $(i \in S)$ such that*

$$\max_{A\in\mathcal{H}} \left| \sum_{i\in A} (\varepsilon_i - p_i) \right| < 2\Delta(\mathcal{H}) .$$

Proof. The key idea is to consider variables ε_i $(i \in S)$ lying anywhere in $[-1, +1]$. Initially $\varepsilon_i = p_i$; all sets then have zero "discrepancy". At the end each ε_i must be -1 or $+1$, providing the coloration in the theorem. We describe the procedure that is to be iterated to go from the initial trivial "coloration" to the final one.

Suppose we have some current assignment ε_i. Call *i fixed* if $\varepsilon_i = \pm 1$ and *floating* otherwise. Let $A = [a_{ij}]$ denote the incidence matrix of the family $\mathcal{H}$. Call row j *ignored* if $\sum' a_{ji} \leqslant \Delta(\mathcal{H})$ (the sum over the floating i) and *active* otherwise. As each column sum is at most $\Delta(\mathcal{H})$, there are fewer active rows than floating columns. Find y_i, for each floating i, with $\sum a_{ji} y_i = 0$ for each active row j. As this system is undetermined, there is a nonzero solution. Now replace ε_i by $\varepsilon_i + \lambda y_i$ where λ is chosen so that all ε_i remain in $[-1, +1]$ and some floating ε_i becomes ± 1 (i.e., fixed).

Iterate the above procedure until all $\varepsilon_i = \pm 1$. To see that the values obtained satisfy the requirement of the theorem, observe that a given row has zero "discrepancy" (i.e., $\sum a_{ji}(\varepsilon_i - p_i) = 0$) until it becomes ignored. After that, each ε_i still floating changes by at most 2 and hence the sum $\sum a_{ji}(\varepsilon_i' - p_i)$ changes by less than $2\Delta(\mathcal{H})$. □

Theorem 2.2 was motivated by the following "integer making lemma" (and in fact is a generalization of it).

Lemma 2.3 (Baranyai 1974). *Let $A = (a_{ij})$ be a matrix of real elements. Then there* *exist an integer matrix* $A^* = (a^*_{ij})$ *such that*

$$|a_{ij} - a^*_{ij}| < 1 \qquad \textit{for all } i, j ,$$

$$\left|\sum_i a_{ij} - \sum_i a^*_{ij}\right| < 1 \quad \textit{for all } j ,$$

$$\left|\sum_j a_{ij} - \sum_j a^*_{ij}\right| < 1 \quad \textit{for all } i ,$$

and

$$\left|\sum_i \sum_j a_{ij} - \sum_i \sum_j a^*_{ij}\right| < 1 .$$

This lemma was the basic tool in Baranyai's theorem on the factorization of the complete uniform hypergraph (see chapters 7 and 14). The lemma can also be proved using the integrality theorem of flow theory (see chapter 2).

We suspect that Theorem 2.2 can be essentially improved. The following conjecture would also generalize Spencer's theorem 2.1.

Conjecture 2.4 (Beck–Fiala).

$$\mathcal{D}(\mathcal{H}) \leq c(\Delta(\mathcal{H}))^{1/2} .$$

If true then it is best possible apart from the constant factor c. Corollary 2.6 below justifies the weaker conjecture $\mathcal{D}(\mathcal{H}) < (\Delta(\mathcal{H}))^{1/2+\varepsilon}$ when both $|S|$ and $|\mathcal{H}|$ are "subexponential" functions of the maximum degree. For later application, we state first a more general result.

Theorem 2.5 (Beck 1981b). *Let* $\mathcal{H}$ *be a finite hypergraph with* $\bigcup \mathcal{H} = S$. *Let* M *and* K *be natural numbers such that*

$$\Delta(\{A \in \mathcal{H}: |A| \geq M\}) \leq K .$$

Then

$$\mathcal{D}(\mathcal{H}) < c(M + K \cdot \log K)^{1/2} \cdot (\log|\mathcal{H}|)^{1/2} \cdot \log|S| .$$

Choosing $M = 1$ and $K = \Delta(H)$, we obtain the following.

Corollary 2.6. *For any finite hypergraph with* $\Delta = \Delta(\mathcal{H})$, *we have*

$$\mathcal{D}(\mathcal{H}) < c \cdot \Delta^{1/2} \cdot \log|\mathcal{H}| \cdot \log|S| .$$

The following somewhat technical theorem, which is useful in applications, is a generalization of Corollary 2.6.

Theorem 2.7 (Beck 1988). *Let $\mathcal{H}$ be a finite hypergraph with $\bigcup\mathcal{H} = S$. Suppose that there is a second family $\mathcal{G}$ of subsets of S such that*

(i) $\Delta(\mathcal{G}) \leqslant D$; *and*

(ii) *every $A \in \mathcal{H}$ can be represented as the disjoint union of at most K elements of $\mathcal{G}$. Then*

$$\mathcal{D}(\mathcal{H}) < c \cdot ((K \cdot D \cdot \log D \cdot \log|\mathcal{H}|)^{1/2} \cdot \log|S| .$$

Note that if $\mathcal{G} = \mathcal{H}$ then we obtain Corollary 2.6.

We have to remark that there are very few general *lower bounds* on $\mathcal{D}(\mathcal{H})$. The following one is based on linear algebra. To state it in its natural generality, define the ℓ_2-discrepancy of a hypergraph $\mathcal{H}$ by

$$\mathcal{D}_2(\mathcal{H}) = \min_{\varepsilon \in \{-1,1\}^S} \Big(\sum_{A \in \mathcal{H}} \Big(\sum_{i \in A} \varepsilon_i\Big)^2\Big)^{1/2} .$$

Clearly $\mathcal{D}(\mathcal{H})m^{-1/2} \leqslant D(\mathcal{H}) \leqslant \mathcal{D}_2(\mathcal{H})$. We denote by $\lambda_{\min}(M)$ the least eigenvalue of the matrix M. We recall: $|\mathcal{H}| = m$ and $|S| = n$.

Theorem 2.8 (Lovász–T. Sós). *Let M be the incidence matrix of $\mathcal{H}$. Then*

(i) $\mathcal{D}_2(\mathcal{H}) \geqslant (n\lambda_{\min}(M^{\mathrm{T}}M))^{1/2}$,

(ii) *if for some diagonal matrix D, the matrix $M^{\mathrm{T}}M - D$ is positive semidefinite, then* $\mathcal{D}_2(\mathcal{H}) \geqslant (\operatorname{Tr} D)^{1/2}$. Note that T stands for transpose.

Proof. Let $f \in \{-1, 1\}^S$ attain the minimum in the definition of $\mathcal{D}_2(\mathcal{H})$. Then

$$\begin{aligned}\mathcal{D}_2(\mathcal{H})^2 &= \sum_{A \in H} \Big(\sum_{i \in A} f_i\Big)^2 = (Mf)^{\mathrm{T}}(Mf) = f^{\mathrm{T}}M^{\mathrm{T}}Mf \\ &\geqslant f^{\mathrm{T}}f\lambda_{\min}(M^{\mathrm{T}}M) = n\lambda_{\min}(M^{\mathrm{T}}M) .\end{aligned}$$

This proves (i); the proof of (ii) is similar. □

Corollary 2.9. *If $\mathcal{H}$ has constant pair-degree, i.e.,*

$$|\{A : i, j \in A \in \mathcal{H}\}| = \lambda$$

for every i, $j \in S$, $i \neq j$, and d_i denotes the degree of $i \in S$, then

$$\mathcal{D}(\mathcal{H}) \geqslant n^{-1/2}\Big(\sum_{i=1}^{n} (d_i - \lambda)\Big)^{1/2} .$$

Corollary 2.10. *Let $\mathcal{H}$ be formed by the set of lines in a finite projective plane of order p. Then*

$$\mathcal{D}(\mathcal{H}) \geqslant \sqrt{p} .$$

Corollary 2.11. *Let H be an $n \times n$ Hadamard matrix, i.e., a ± 1 matrix whose column vectors are mutually orthogonal and has all 1s in the first row. Let $\mathcal{H}$ be the* *hypergraph whose incidence matrix is obtained from H by replacing the* -1s *by* 0s. *Then*

$$\mathcal{D}(\mathcal{H}) > \frac{\sqrt{n}}{2}.$$

This corollary proves that Theorem 2.1 is best possible apart from the constant factor.

The most important application of Theorem 2.8 is Roth's theorem (Theorem 1.2, see section 5).

3. Various concepts of discrepancy

Suppose we want to split the sets in $\mathcal{H}$ in ratio α, $1-\alpha$. In other words, we want to find a system of representatives of $\mathcal{H}$ so that the number of representatives in every set $A \in \mathcal{H}$ is as close to $\alpha|A|$ as possible. Then, setting $\lambda = 2\alpha - 1$,

$$\mathcal{D}(\mathcal{H}; \lambda) = \min_{\varepsilon \in \{-1,1\}^S} \max_{A \in \mathcal{H}} \left| \sum_{i \in A} (\varepsilon_i - \lambda) \right|$$

measures the corresponding discrepancy. Obviously

$$\mathcal{D}(\mathcal{H}; \tfrac{1}{2}) = \mathcal{D}(\mathcal{H}).$$

More generally, we may consider a weight-function $p: S \to [-1, 1]$ and the corresponding discrepancy

$$\mathcal{D}(\mathcal{H}; p) = \min_{\varepsilon \in \{-1,1\}^S} \max_{A \in \mathcal{H}} \left| \sum_{i \in A} (\varepsilon_i - p_i) \right|$$

(this value has come up in Theorem 2.2′). The *inhomogeneous discrepancy* of $\mathcal{H}$ is defined by

$$\mathcal{D}_1(\mathcal{H}) = \max_p \mathcal{D}(\mathcal{H}, p)$$

and measures how well an arbitrary weight distribution on S can be approximated with 0–1 measures regarding the family $\mathcal{H}$. Considering the particular cases $p_1 = \cdots = p_n = \lambda$ we define the *diagonal discrepancy* by

$$\mathcal{D}_{\mathrm{D}}(\mathcal{H}) = \max_\lambda \mathcal{D}(\mathcal{H}; \lambda).$$

The *hereditary discrepancy* of $\mathcal{H}$ is defined by

$$\mathcal{D}_{\mathrm{H}}(\mathcal{H}) = \sup_{X \subseteq S} \mathcal{D}(\mathcal{H}_X).$$

Ghouila-Houri's theorem 1.5 asserts that a hypergraph is totally unimodular iff its hereditary discrepancy is at most 1.

Observe that adding new elements to some of the sets in $\mathcal{H}$ appropriately we can achieve that this enlarged hypergraph will have discrepancy 0. This means, that $\mathcal{D}(\mathcal{H})$ can be small by accident, while $\mathcal{D}_I(\mathcal{H})$ and $\mathcal{D}_H(\mathcal{H})$ depend on more intrinsic properties of $\mathcal{H}$. In fact, $\mathcal{D}(\mathcal{H})$ can be much smaller then $\mathcal{D}_I(\mathcal{H})$ or $\mathcal{D}_H(\mathcal{H})$. A simple example is the following. Let $S = \{1, \dots, 4n\}$ and

$$\mathcal{H} = \{A \mid A \subset S, |A \cap \{1, \dots, 2n\}| = |A|/2\} .$$

Then $\mathcal{D}(\mathcal{H}) = 0$ but $\mathcal{D}_I(\mathcal{H}) = n$ and $\mathcal{D}_H(\mathcal{H}) = n$.

We mention the trivial inequalities

$$\mathcal{D}(\mathcal{H}) \leqslant \min\{\mathcal{D}_I(\mathcal{H}), \mathcal{D}_H(\mathcal{H})\}$$

and

$$\mathcal{D}(\mathcal{H}, \lambda) \leqslant \mathcal{D}_D(\mathcal{H}) \leqslant \mathcal{D}_I(\mathcal{H}) .$$

The following nontrivial inequality was first explicitly formulated in Lovász et al. (1986). The proof is identical with that of Lemma 3 in Beck and Spencer (1984b).

Theorem 3.1. *For every hypergraph* $\mathcal{H}$,

$$\mathcal{D}_I(\mathcal{H}) \leqslant 2\mathcal{D}_H(\mathcal{H}) .$$

Proof. Let, for each $i \in S$, a weight $-1 \leqslant p_i \leqslant 1$ be given. Let $\alpha_i = (1 + p_i)/2 \in [0, 1]$. Assume first that all the α_i have finite binary expansion, i.e., there is a natural number n so that $2^n \cdot \alpha_i \in \mathbb{Z}$ for all $i \in S$. Let n be minimal with this property. Let $X \subset S$ be the set of points $i \in S$ such that α_i has 1 for its nth binary digit. As $\mathcal{D}(\mathcal{H}) \leqslant \mathcal{D}_H(\mathcal{H})$, there exist $\varepsilon_i = \pm 1$ for all $i \in X$ such that

$$\Big|\sum_{i \in A \cap X} \varepsilon_i\Big| \leqslant \mathcal{D}_H(\mathcal{H})$$

for all $A \in \mathcal{H}$. Define approximations $\alpha_1^{(1)}, \alpha_2^{(1)}, \dots, \alpha_N^{(1)}$ by

$$\alpha_i^{(1)} = \begin{cases} \alpha_i + \varepsilon_i \cdot 2^{-n} & \text{if } i \in X , \\ \alpha_i & \text{if } i \in S \setminus X . \end{cases}$$

For any $A \in \mathcal{H}$,

$$\Big|\sum_{i \in A} (\alpha_i^{(1)} - \alpha_i)\Big| = \Big|\sum_{i \in A \cap X} 2^{-n} \cdot \varepsilon_i\Big| \leqslant 2^{-n} \cdot \mathcal{D}_H(\mathcal{H}) .$$

The values $\alpha_i^{(1)}$ have binary expansions of length at most $(n-1)$. We repeat this procedure (note that X will be a different set), getting $\alpha_i^{(2)}$ with

$$\Big|\sum_{i \in A} (\alpha_i^{(2)} - \alpha_i^{(1)})\Big| \leqslant 2^{-(n-1)} \mathcal{D}_H(\mathcal{H})$$

for all $A \in \mathcal{H}$.

We apply this procedure n times, finally reaching $\alpha_i^{(n)}$ with binary expansions of length zero, i.e., $\alpha_i^{(n)}=0$ or 1. Let $\varepsilon_i=2\alpha_i^{(n)}-1\in\{-1,+1\}$. Then for all $A\in\mathcal{H}$,

$$\left|\sum_{x_i\in A}(\varepsilon_i-p_i)\right|=2\left|\sum_{i\in A}(\alpha_i^{(n)}-\alpha_i)\right|\leqslant 2\sum_{j=0}^{n-1}\left|\sum_{i\in A}(\alpha_i^{(j+1)}-\alpha_i^{(j)})\right|$$
$$\leqslant 2\sum_{j=0}^{n-1}2^{-(n-j)}\cdot\mathcal{D}_{\mathrm{H}}(\mathcal{H})\leqslant 2\mathcal{D}_{\mathrm{H}}(\mathcal{H})$$

as required. Finally, a compactness argument implies the truth of Theorem 3.1 for arbitrary $p_1,\ldots,p_n\in[-1,+1]$. □

Observe that all the upper bounds in Theorems 2.1, 2.2, 2.6, 2.7 are valid in fact for the hereditary discrepancy.

The discrepancy of a matrix. The concept of discrepancy can be expressed in terms of the incidence matrix M of the hypergraph $\mathcal{H}$:

$$\mathcal{D}(\mathcal{H})=\min_{\varepsilon\in\{-1,+1\}^S}\|M\varepsilon\|_\infty$$

and

$$\mathcal{D}_{\mathrm{I}}(\mathcal{H})=\max_{p\in[-1,+1]^S}\ \min_{\varepsilon\in\{-1,+1\}^S}\|M(\varepsilon-p)\|_\infty\,.$$

Note that these definitions are meaningful for any matrix M. Therefore, following Lovász et al. (1986), we can use the notation $\mathcal{D}(M)$ and $\mathcal{D}_{\mathrm{I}}(M)$ for an arbitrary matrix M. We can also generalize the hereditary version by letting $\mathcal{D}_{\mathrm{H}}(M)$ be the maximum of $\mathcal{D}(M')$ over all submatrices M' of M.

Almost all of the previous results, most notably Theorems 2.2 and 2.8, extend to matrices in a natural way. The following slight generalization of Theorem 2.2 also follows by the same argument.

Theorem 3.2. *Assume that every square submatrix of a matrix M has row with l_1-norm at most* 1. *Then $\mathcal{D}(M)\leqslant 2$.*

The above generalized versions of the notion of discrepancy may become easier to grasp from the following nice geometric interpretation. Consider the set

$$U_A=\{x\in R^S:\ \|Ax\|_\infty\leqslant 1\}\ ,$$

i.e., the "unit ball" of the norm $\|Ax\|_\infty$. So U_A is a convex polyhedron centrally symmetric with respect to the origin. For $t>0$, consider the convex set $t\cdot U_A$ and let $U_1(t), U_2(t),\ldots$ be the copies of $t\cdot U_a$ obtained by translating its center by all ±1-vectors. Then

- $\mathcal{D}(A)$ is the least number t for which some $U_j(t)$ contains the origin;
- $\mathcal{D}_l(A)$ is the least number t for which the sets $U_j(t)$ cover the cube $[-1,1]^S$;

- $\mathcal{D}(A)$ is the least number t for which the center of each face F of the cube $[-1, 1]^S$ is contained in at least one of the sets $U_j(t)$ centered at the vertices of F.

Theorem 1.5 raises the question whether in general the discrepancy of a hypergraph (or of a matrix) is related to the determinants of the submatrices of the incidence matrix. In this direction there is a lower bound theorem from Lovász et al. (1986).

Theorem 3.3. *For any matrix A,*

$$\mathcal{D}(A) \geq \max_k \max_B |\det B|^{1/k} ,$$

where B ranges over all $k \times k$ submatrices of A.

Let us think of the rows of matrix A as ordered by *importance* so that we may wish to make the discrepancy in early rows extremely small, perhaps at the expense of the later E_i. The following result states that there is an approximation which is extremely good with respect to the early rows and is reasonably good with respect to all.

Theorem 3.4 (Beck and Spencer 1984b, Spencer 1985). *Let $M = (m_{ij}) \in \mathbb{R}^{m \times n}$ be a matrix with $|m_{ij}| \leq 1$. Let $p_1, \ldots, p_N \in [-1, +1]$.*

(i) *There exist $\varepsilon_1, \ldots, \varepsilon_n \in \{-1, +1\}$ so that*

$$\left| \sum_{j=1}^{n} m_{ij}(p_i - \varepsilon_i) \right| < ci^{1/2} ,$$

(ii) *If the upper bound is relaxed to $2i$ then such ε_j are polynomial time computable.*

Note that (i) of Theorem 3.4 is best possible apart from constant factor (this again follows by considering Hadamard matrices). Part (ii) follows by applying Theorem 3.3 (whose proof, just like the proof of Theorem 2.2, can be followed by a polynomial time algorithm) to the matrix (m_{ij}/i).

In the particular case $m_{ij} \in \{0, 1\}$ and $p_i = 0$ we obtain the following.

Corollary 3.5. *Let $Y_1, Y_2, Y_3, \ldots, Y_M$ be a sequence of subsets of a finite set X.*

(i) *There exist a 2-coloring $f: X \to \{-1, +1\}$ so that*

$$\left| \sum_{x \in Y_i} f(x) \right| < c \cdot i^{1/2} , \quad 1 \leq i \leq M .$$

(ii) *One can find in polynomial time a 2-coloring $f: X \to \{-1, +1\}$ so that*

$$\left| \sum_{x \in Y_i} f(x) \right| < 2i , \quad 1 \leq i \leq M .$$

Theorem 3.4 has some nice applications in a matrix balancing problem (see

Beck and Spencer 1983, 1989). Let an arbitrary matrix $A=(a_{ij})$, $1\leq i,\ j\leq n$, be given with all $a_{ij}\in\{-1,+1\}$. By a row shift we mean the act of replacing, for a particular i, all coefficients a_{ij} in the ith row by their negatives $(-a_{ij})$. The column shift is defined similarly. A line shift means either a row or a column shift. Consider the following solitaire game. The player applies a succession of line shifts to the matrix A. His object is to make the absolute value of the sum of all the coefficients of A as small as possible. Let $\|A\|$ denote this minimum value. Komlós and Sulyok (1970), resolving a conjecture of L. Moser, showed that if n is sufficiently large then $\|A\|\leq 2$ may be achieved ($\|A\|\leq 1$ if n is odd). As an illustration, we shall derive this result from (ii) of Theorem 3.4 in the case of even n.

Theorem 3.6. *Let $n\geq 2$ be an even integer. Given any $n\times n$ matrix $A=(a_{ij})$ with all $a_{ij}\in\{-1,+1\}$, there exist $\delta_1,\ldots,\delta_n$, $\varepsilon_1,\ldots,\varepsilon_n\in\{-1,+1\}$ so that*

$$\|A\|=\left|\sum_{i=1}^{n}\sum_{j=1}^{n}\delta_i\varepsilon_j a_{ij}\right|\leq 2\,.$$

Proof. It follows from (ii) of Theorem 3.4 that there exist column shifts ε_j so that the new row sums r_i satisfy $|r_i|<2i$, $1\leq i\leq K$. For simplicity of notation let us then apply row shifts so that all row sums are nonnegative. Since all r_i are even integers we have $r_1=0$, and, in general, $0\leq r_i\leq 2i-2$.

We now describe a simple technique that will give the final row shifts. Let $s_1,\ldots,s_K$ be nonnegative integers and let T be a positive integer such that $s_1\leq T$ and for $1\leq i\leq n-1$,

$$s_{i+1}\leq s_1+\cdots+s_i+T\,.$$

Then there exist $\delta_i,\ldots,\delta_n=\pm 1$, so that

$$|\delta_1 s_1+\cdots+\delta_n s_n|\leq T\,.$$

We can find such δ_i by reverse induction. Set $\delta_n=+1$. Having found δ_n, $\delta_{n-1},\ldots,\delta_{i+1}$ we choose $\delta_i=\pm 1$ so as to minimize the absolute value of the partial sum $\delta_n s_n+\cdots+\delta_{i+1}s_{i+1}+\delta_i s_i$. We shall call this method the *greedy technique* for the remainder of the proof.

We may not immediately apply the greedy technique because we may have too many $r_i=0$ and thereby T large. Reorder the rows in increasing order of row sums. We then still have $0\leq r_i\leq 2i-2$. Suppose the first u rows have sum zero and the next v rows have sum two. If $u=1$ we may simply apply the greedy technique so we shall assume $u>1$. Let r_i' be the new absolute value of ith row sum after a single column is shifted. For the first u rows $r_i'=2$ regardless of which column is shifted. For the next v rows $r_i'=0$ for $(n/2)+1$ of the possible column shifts, these being the cases when an entry $+1$ switched to -1, and $r_i'=4$ for the remaining $(n/2)-1$ column shifts. Thus the average value of r_i', taken over all n possible column shifts, is $2-(4/n)$. Now we conclude that the average value of $r_{u+1}'+\cdots+r_{u+v}'$ is $v(2-(4/n))$. If $v\geq n/2$ then the greedy technique trivially works and hence we may assume $v < n/2$. Thus $v(2-(4/n)) > 2v - 2$. Since this is the average, there must be one specific column change so that

(1) $r'_{u+1} + \cdots + r'_{u+v} \geq 2v$. We also have

(2) $r'_1 = \cdots = r'_u = 2$, and

(3) $0 \leq r'_i \leq 2i$ for $i > u + v$.

We observe that $r'_1 + \cdots + r'_{u+v} \geq 2(u+v)$ and $r'_i \geq 2$ for $i > u+v$ since $r_i \geq 4$. Hence

$$r'_1 + \cdots + r'_i + 2 \geq 2i + 2 \geq r'_{i+1}$$

for all $i \geq u + v$.

Trivially

$$r'_1 + \cdots + r'_i + 2 \geq 4 \geq r'_{i+1} ,$$

when $1 \leq i < u + v$. Thus we may apply the greedy technique to the row sums $r'_1, \ldots, r'_k$, completing the proof. □

Applying the stronger relation (i) of Theorem 3.3, one can prove the following general result (see Beck and Spencer 1989).

Theorem 3.7. *There exists a constant $c > 0$ such that for every $m \times n$ matrix $A = (a_{ij})$ with all $|a_{ij}| \leq 1$ there exist $\delta \in \{-1, 1\}^m$ and $\varepsilon \in \{-1, 1\}^n$ such that*

$$\left|\sum_i \sum_j \delta_i \varepsilon_j a_{ij}\right| < c .$$

4. Vector-sums

We have seen a geometric interpretation of discrepancy problems in the row space of the corresponding matrix. Now we consider the space of the column vectors, which leads to several new and interesting questions. In fact the investigation of value-distributions of *vector-sums* developed earlier and independently of hypergraph coloring problems or of discrepancy theory.

Let $M = (v_1, \ldots, v_n)$, $v_i \in \mathbb{R}^m$ for $1 \leq i \leq n$. Let further $\|\cdot\|$ and $\|\cdot\|'$ denote two arbitrary norms in $\mathbb{R}^m$.

We define the *discrepancy* (relative to the two norms) by

$$\mathcal{D}(M; \|\cdot\|, \|\cdot\|') = \frac{\min_{\varepsilon \in \{-1,1\}^n} \|\sum_{i=1}^n \varepsilon_i v_i\|}{\max_{1 \leq i \leq n} \|v_i\|'}$$

and

$$\mathcal{D}(\|\cdot\|, \|\cdot\|') = \max_M \mathcal{D}(M; \|\cdot\|, \|\cdot\|') .$$

Note that for any matrix M and norm $\|\cdot\|'$,

$$\mathscr{D}(M)=\mathscr{D}(M;l_\infty,\|\cdot\|')\cdot\max_{1\leqslant i\leqslant n}\|v_i\|'\ .$$

The case when $\|\cdot\|$ was the l_2 norm also came up briefly in Theorem 2.8.

Already in 1963 it was asked by Dworetzky what $\mathscr{D}(\|\cdot\|,\|\cdot\|)$ equals for a given norm. The more general question (where the two norms are not necessarily the same) was formulated first in Bárány and Grinberg (1981), who gave the following general upper bound for Dworetzky's problem.

Theorem 4.1 (Bárány and Grinberg 1981). *For an arbitrary norm $\|\cdot\|$ in $\mathbb{R}^m$,*

$$\mathscr{D}(\|\cdot\|,\|\cdot\|)\leqslant m\ .$$

This is sharp when $\|\cdot\|$ is the l_1 norm.

Now let us consider the special cases when $\|\cdot\|$ and $\|\cdot\|^*$ are one of the three most important norms: the l_∞ norm, the l_2 norm or the l_1 norm. Theorem 2.1 has the following generalization in this setting.

Theorem 4.2 (Spencer 1985). $\mathscr{D}(l_\infty,l_\infty)\leqslant 6\sqrt{m}$.

(Observe that the upper bounds in Theorems 4.1 and 4.2 depend only on the dimension!) Theorem 2.2 is also valid in this more general form.

Theorem 4.3 (Beck and Fiala 1981). $\mathscr{D}(l_\infty,l_1)\leqslant 2$.

Grinberg observed, that for any M in $\mathbb{R}^m$,

$$\mathscr{D}(l_2,l_2)\leqslant\sqrt{m}\ .$$

This is sharp. Indeed, consider m pairwise orthogonal unit vectors $e_1,\ldots,e_m$ in $\mathbb{R}^m$. Then $\|\sum_{i=1}^m\varepsilon_ie_i\|_2=m^{1/2}$ for any choice of $\varepsilon\in\{-1,1\}^m$.

All but one of the remaining cases are trivial or easy consequences of the above ones. The only nontrivial case is when $\|\cdot\|=l_\infty$ and $\|\cdot\|'=l_2$. In that case nothing nontrivial is known. The conjecture of Komlós refers to this case.

Conjecture 4.4 (Komlós). There exists an absolute constant c such that

$$\mathscr{D}(l_\infty,l_2)\leqslant c\ .$$

The Komlós conjecture implies the Beck–Fiala conjecture 2.4 for set-systems.

Partial sums. The problems we considered in the preceding paragraphs are of *static character*. The *dynamic* version is when we color the points one by one and we would like to have a "good" coloring at *each stage*. This formulation also allows us to study problems in which the underlying set S is infinite. The following theorems are of this "dynamic" character.

Theorem 4.5 (Bárány and Grinberg 1981). *Let $v_1, v_2, \ldots, v_n$ be n vectors in $\mathbb{R}^m$ with $\|v_i\| \leqslant 1$, where $\|\cdot\|$ is any norm in $\mathbb{R}^m$. Then there exist a sequence $\varepsilon_1, \ldots, \varepsilon_n$, $\varepsilon_i \in \{-1, +1\}$ so that*

$$\Big\|\sum_{i=1}^{t} \varepsilon_i v_i\Big\| \leqslant 2m\,, \quad \textit{for } t = 1, 2, \ldots, n\,.$$

It is conjectured that if $\|\cdot\| = l_2$ or l_∞, then in this theorem, $2m$ can be replaced by $K\sqrt{m}$. For l_∞ norm and if $m = n$, Spencer proved this conjecture.

Theorem 4.6 (Spencer 1986). *For any sequence $v_1, \ldots, v_n$ of vectors in $\mathbb{R}^m$ with $\|v_i\|_\infty \leqslant 1$, there exists a sequence $\varepsilon_1, \ldots, \varepsilon_n$, $\varepsilon_i \in \{+1, -1\}$ so that*

$$\Big\|\sum_{i=1}^{t} \varepsilon_i v_i\Big\|_\infty \leqslant K\sqrt{m} \quad \textit{for } t = 1, \ldots, n\,.$$

An infinite-dimensional version of Theorem 4.5 is the following.

Theorem 4.7 (Beck 1990). *Let $v_1, v_2, v_3, \ldots$ be infinite-dimensional vectors satisfying $\|v_i\|_\infty \leqslant 1$. Then there exist $\varepsilon_1, \varepsilon_2, \varepsilon_3, \ldots$; $\varepsilon_i \in \{-1, +1\}$ so that*

$$\Big|\Big(\sum_{i=1}^{t} \varepsilon_i v_i\Big)_j\Big| \leqslant j^{4+o(1)}$$

for all j and t. Here v_j stands for the jth coordinate of the vector v.

Permutation of vectors. Instead of flipping the sign of vectors, we may achieve that all partial sums be small just by rearranging them. In fact, the two kinds of problems are strongly related as the following "transference lemma" of Chobayan shows.

Theorem 4.8. *Let $v_1, \ldots, v_n \in \mathbb{R}^m$ with $v_1 + v_2 + \cdots + v_n = 0$, and let $\|\cdot\|$ be an arbitrary norm in $\mathbb{R}^m$. Suppose that for every permutation $\pi = (i_1, i_2, \ldots, i_n)$ of $\{1, 2, \ldots, n\}$ there exist $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n \in \{-1, +1\}$ (depending on π) such that*

$$\max_{1 \leqslant t \leqslant n} \Big\|\sum_{j=1}^{t} \varepsilon_j v_{i_j}\Big\| \leqslant A\,.$$

Then there is a permutation $\pi^* = (l_1, l_2, \ldots, l_n)$ of $\{1, 2, \ldots, n\}$ such that

$$\max_{1 \leqslant t \leqslant n} \Big\|\sum_{j=1}^{t} v_{l_j}\Big\| \leqslant A\,.$$

Proof. Let

$$B = \min_{\pi=(i_1,\ldots,i_n)} \max_{1\leq t\leq n} \left\| \sum_{j=1}^{t} v_{i_j} \right\| .$$

We have to show that $B \leq A$. Let $\pi^* = (l_1, l_2, \ldots, l_n)$ denote the permutation where the minimum is attained. By the hypothesis of the theorem, there exist $\varepsilon_1^*, \ldots, \varepsilon_n^* \in \{-1, +1\}$ such that

$$\left\| \sum_{j=1}^{t} \varepsilon_j^* v_{l_j} \right\| \leq A \quad \text{for all } 1 \leq t \leq n .$$

Let

$$M^+ = \{1 \leq j \leq n : \varepsilon_j^* = +1\} , \qquad M^- = \{1 \leq j \leq n : \varepsilon_j^* = -1\} .$$

We have

$$\sum_{j=1}^{t} v_{l_j} + \sum_{j=1}^{t} \varepsilon_j^* v_{l_j} = 2 \sum_{\substack{j\in M^+\\ 1\leq j\leq t}} v_{l_j} ,$$

and

$$\sum_{j=1}^{t} v_{l_j} - \sum_{j=1}^{t} \varepsilon_j^* v_{l_j} = 2 \sum_{\substack{j\in M^-\\ 1\leq j\leq t}} v_{l_j} .$$

Hence

$$\left\| \sum_{\substack{j\in M^+\\ 1\leq j\leq t}} v_{l_j} \right\| \leq \frac{A+B}{2} ,$$

and

$$\left\| \sum_{\substack{j\in M^-\\ 1\leq j\leq t}} v_{l_j} \right\| \leq \frac{A+B}{2} .$$

Setting $M^+ = \{p_1 < p_2 < \cdots < p_r\}$ and $M^- = \{q_1 < q_2 < \cdots < q_s\}$, we define the permutation

$$\pi^{**} = (p_1, p_2, \ldots, p_r, q_s, q_{s-1}, \ldots, q_2, q_1) ,$$

which we also denote by $(h_1, \ldots, h_n)$. It follows from the assumption $v_1 + v_2 + \cdots + v_n = 0$ that

$$\max_{1\leq t\leq n} \left\| \sum_{j=1}^{t} v_{h_j} \right\| \leq \frac{A+B}{2} .$$

Since B was the minimum, we must have $B \leqslant (A+B)/2$, and the desired inequality $B \leqslant A$ follows. □

Combining Theorems 4.5 and 4.7 with Chobanyan's transference lemma, we get the following result.

Corollary 4.9 (Bárány and Grinberg 1981). *Let $v_1, \ldots, v_n$ be n vectors in $\mathbb{R}^m$ with $\|v_i\| \leqslant 1$ where $\|\cdot\|$ is any norm in $\mathbb{R}^m$. Assume that $v_1 + v_2 + \cdots + v_n = 0$. Then there exists a permutation $v_{i_1}, v_{i_2}, \ldots, v_{i_n}$ of the vectors v_i such that*

$$\max_{1 \leqslant t \leqslant n} \left\| \sum_{j=1}^{t} v_{i_j} \right\| \leqslant 2m .$$

Corollary 4.10. *Let $v_1, \ldots, v_n$ be infinite-dimensional vectors satisfying $\|v_i\|_\infty \leqslant 1$ $(1 \leqslant i \leqslant n)$ and $v_1 + v_2 + \cdots + v_n = 0$. Then there exists a permutation $v_{i_1}, v_{i_2}, \ldots, v_{i_n}$ of the vectors v_i such that*

$$\left| \max_{1 \leqslant t \leqslant n} \left(\sum_{j=1}^{t} v_{i_j} \right)_l \right| \leqslant l^{4+o(1)}$$

for all $l \geqslant 1$.

5. Arithmetic progressions

A structure whose discrepancy properties have been extensively investigated is the family of arithmetic progressions. We have seen Roth's theorem 1.2 and Van der Waerden's theorem 1.3, showing the two sides of (qualitatively) the same phenomenon: If we focus on the short arithmetic progressions, we get a monochromatic one; if we focus on longer arithmetic progressions, a weaker preponderance phenomenon (large discrepancy) can be asserted.

Van der Waerden's theorem and related results on arithmetic progressions are discussed in chapter 25. Here we treat the ramifications of Roth's theorem 1.2. Let us reformulate it in the language introduced above. Let $\mathcal{H}_n$ denote the hypergraph formed by the arithmetic progressions in $\{1, \ldots, n\}$. Then

Theorem 5.1. $\mathcal{D}(\mathcal{H}_n) > cn^{1/4}$.

Proof. Let $k = \lfloor \sqrt{n/6} \rfloor$. We show that the arithmetic progression can be chosen of length k and of difference at most $6k$. Let us allow, however, also "wrapped" arithmetic progressions, i.e., subsets of $\{1, \ldots, n\}$ that arise from an arithmetic progression of length k and difference at most $6k$ by reduction modulo n. By the choice of k, every "wrapped" progression is the union of two "proper" arithmetic progressions, and hence it suffices to prove that if $\mathcal{H}$ is the hypergraph formed by "wrapped" arithmetic progressions, then $\mathcal{H}$ has discrepancy at least $\frac{1}{10}n^{1/4}$. Note that $m = |\mathcal{H}| = 6kn$.

Let M be the incidence matrix of $\mathcal{H}$. By Theorem 2.8(i), it suffices estimate $\lambda_{\min}(M^{\mathrm{T}}M)$ from below.

Now the matrix $M^{\mathrm{T}}M$ is a circulant (this is where wrapping is needed!), and hence we know that its eigenvectors are $(1, \epsilon, \epsilon^2, \ldots, \epsilon^{n-1})^{\mathrm{T}}$, where ϵ is an nth root of unity. The corresponding eigenvalues are

$$\lambda(\epsilon) = \frac{1}{n} \sum_{A \in \mathcal{H}} \left| \sum_{j \in A} \epsilon^j \right|^2 .$$

Note that for each arithmetic progression A, there are $n - 1$ others (its translates) that give the same contribution. So we may just select arithmetic progressions starting at 0:

$$\lambda(\epsilon) = \sum_{d=1}^{6k} \left| \sum_{t=0}^{k-1} \epsilon^{td} \right|^2 .$$

By the pigeon hole principle we can find a d_0, $1 \leqslant d_0 \leqslant k$ such that

$$-\pi/(3k) \leqslant \arg(\epsilon^{d_0}) \leqslant \pi/(3k) .$$

Then $\operatorname{Re} \epsilon^{td_0} \geqslant \frac{1}{2}$ for $1 \leqslant t \leqslant k - 1$, and hence

$$\lambda(\epsilon) \geqslant \left| \sum_{t=0}^{k-1} \epsilon^{td_0} \right|^2 \geqslant \left(\operatorname{Re} \sum_{t=0}^{k-1} \epsilon^{td} \right)^2 \geqslant \frac{k^2}{4} .$$

Thus

$$\mathcal{D}(\mathcal{H}) \geqslant \frac{1}{\sqrt{m}} \mathcal{D}_2(\mathcal{H}) \geqslant \left(\frac{n}{m} \lambda_{\min} \right)^{1/2} \geqslant \left(\frac{k}{24} \right)^{1/2} > \tfrac{1}{10} n^{1/4} . \qquad \square$$

Note that we have actually proved a stronger, l_2 norm version. This gives the following information about the difference d of the arithmetic progressions of large discrepancy.

Corollary 5.2 (Roth). *Given any 2-coloring $f : \mathbb{N} \to \{-1, +1\}$ of the natural numbers, for infinitely many values of d, there is an arithmetic progression $P = P(d)$ of difference d such that*

$$\left| \sum_{k \in P(d)} f(k) \right| > c\sqrt{d} .$$

Roth conjectured that the exponent $\frac{1}{4}$ of N in Theorem 1.2 can be improved to $\frac{1}{2}$ (which corresponds to the random 2-coloring). This was disproved by Sárközy (1973). Beck (1981b) proved that Roths's lower bound is nearly sharp, by a combinatorial argument based on Theorem 2.5.

Theorem 5.3.* $\mathcal{D}(\mathcal{H}_n) < c \cdot n^{1/4} \cdot (\log n)^3$.

Proof. For integers satisfying $i \leqslant j$, let

$$\mathrm{AP}(a, d, i, j) = \{a + k \cdot d \colon i \leqslant k \leqslant j\} ,$$

i.e., $\mathrm{AP}(a, d, i, j)$ denotes the arithmetic progression with difference d, starting from $(a + i \cdot d)$ and terminating at $(a + j \cdot d)$. We shall say that an arithmetic progression is special if it is of the type

$$\mathrm{AP}(b, d, i \cdot 2^s, (i + 1) \cdot 2^s - 1) ,$$

where $d \geqslant 1$, $1 \leqslant b \leqslant d$, $i \geqslant 0$ and $s \geqslant 0$. Let $\mathcal{H}_n^*$ denote the family of special arithmetic progressions contained in $\{1, 2, \ldots, n\}$. By definition,

$$\begin{aligned} \Delta(\{A \in \mathcal{H}_n^* \colon |A| \geqslant M\}) &= \max_{1 \leqslant k \leqslant n} |\{A \in \mathcal{H}_n^* \colon |A| \geqslant M \text{ and } k \in A\}| \\ &\leqslant \max_{1 \leqslant k \leqslant n} \sum_{1 \leqslant d \leqslant \frac{n-1}{M-1}} \sum_{\substack{1 < b < d \\ b \equiv k \ (\mathrm{mod}\ d)}} \sum_{\substack{s \\ 2^s \geqslant M \\ b + (2^s - 1)d \leqslant n}} 1 . \end{aligned}$$

Simple calculation shows that the innermost sum is at most $c \cdot \log(n/(d \cdot M))$. It follows that

$$\begin{aligned} \Delta(\{A \in \mathcal{H}_n^* \colon |A| \geqslant M\}) &\leqslant c \cdot \max_{1 \leqslant k \leqslant n} \sum_{1 \leqslant d \leqslant \frac{n-1}{M-1}} \sum_{\substack{1 \leqslant b \leqslant d \\ b \equiv k \ (\mathrm{mod}\ d)}} \log\left(\frac{n}{d \cdot M}\right) \\ &= c \cdot \sum_{1 \leqslant d \leqslant \frac{n-1}{M-1}} \log\left(\frac{n}{d \cdot M}\right) \leqslant c_1 \cdot \frac{n}{M} . \end{aligned}$$

Now we apply Theorem 2.5 to $\mathcal{H}_n^*$ with $M = D = \lceil (c_1 n)^{1/2} \rceil$. Then we obtain

$$\mathcal{D}(\mathcal{H}_n^*) < c_2 \cdot n^{1/4} \cdot (\log n)^2 .$$

We claim that

$$\mathcal{D}(\mathcal{H}_n) \leqslant (2 \log_2 n) \cdot \mathcal{D}(\mathcal{H}_n^*) .$$

To see this, first observe that any arithmetic progression a, $a + d, \ldots, a + l \cdot d$ in $[1, n]$ is representable in the form

$$\mathrm{AP}(b, d, 0, p_1) \backslash \mathrm{AP}(b, d, 0, p_2) ,$$

where $a = b + (p_2 + 1)d$, $1 \leqslant b \leqslant d$ and $p_1 = p_2 + 1 + l$. Moreover, both $\mathrm{AP}(b, d, 0, p_i)$ $(i = 1, 2)$ are disjoint unions of not more than $\log_2 n$ special

* Very recently Matousek and Spencer (1994) cancelled the factor $(\log n)^3$. The new idea is a clever application of a lemma of Haussler. See also Matousek (1994).

arithmetic progressions, i.e., elements of $\mathcal{H}_n^*$. Hence the "best" 2-coloring of $\mathcal{H}_n^*$ gives a 2-coloring of $\mathcal{H}_n$ with discrepancy at most $(2\log_2 n)$ times as large. □

The following result is a sort of converse of Corollary 5.2.

Theorem 5.4 (Beck and Spencer 1984a). *Let n be a positive integer. Then there exists a 2-coloring $f: \mathbb{N} \to \{-1, +1\}$ of the natural numbers such that for any arithmetic progression $P = P(d) = \{a, a+d, a+2d, \ldots\}$ of difference $d \leq n$ and of arbitrary length,*

$$\left|\sum_{k\in P(d)} f(k)\right| < c\cdot\sqrt{d}\cdot(\log n)^{3.5} \quad (1 \leq d \leq n)\,.$$

Unfortunately, we cannot prove that Theorem 5.4 is true with the right-hand side replaced by $d^{(1/2)+o(1)}$. As an upper bound depending only on the difference d of the progression, the weaker estimate $d^{8+o(1)}$ immediately follows from Theorem 4.7.

There is still no answer to the following old conjecture of P. Erdős (worth of $\geq \$500$).

Conjecture 5.5. For any $f: \mathbb{N} \to \{-1, +1\}$ and for every constant C there are a d and n so that

$$\left|\sum_{k=1}^{n} f(k\cdot d)\right| > C\,.$$

In other words, the family of arithmetic progressions with first term 0 has unbounded discrepancy.

6. Measure theoretic discrepancy

We find the roots of discrepancy theory in number theory, in the theory of uniformly distributed sequences, and we give a brief introduction to this theory. (For the general theory of uniformly distributed sequences see the book of Kuipers and Niederreiter 1974.)

The field originated with the celebrated paper of Weyl (1916), which was intended to furnish a deeper understanding of the results in diophantine approximation and to generalize some basic results in this field. At the beginning of this century, due to the work of Ostrowski, Hecke, Hardy, Littlewood, and others, it became clear that the approximability properties of an irrational α by rationals depends on the partial quotients (the "digits" a_k) in its continued fraction expansion

$$\alpha = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cdots}}\,.$$

It became also clear that the approximability property of α is closely related to the distribution of the sequence $(\{n\alpha\})$ in $[0, 1)$. ($\{x\}$ stands for the fractional part of the real number x.)

For every irrational α, the sequence $(\{n\alpha\})$ is everywhere dense in $[0, 1)$. The fact that it is *uniformly distributed* expresses a stronger property. Let us give the definition for arbitrary dimension.

Let $\omega = (u_n)$, $n \in \mathbb{N}$ be a sequence in the k-dimensional unit cube $[0, 1)^k$. Let $B(a, b) = \prod_{i=1}^{k} [a_i, b_i)$ be an aligned box in $[0, 1)^k$, and $\mathcal{B}$, the family of all such boxes. $Z(B; N)$ will denote the number of v with $u_v \in B$, $1 \leqslant v \leqslant N$. Let $R([0, 1]^k)$ denote the set of Riemann-integrable functions on $[0, 1]^k$.

Definition 6.1. The sequence $\omega = (u_n)$, $n \in \mathbb{N}$ is said to be uniformly distributed in $[0, 1)^k$ if for every aligned box $B \subset [0, 1)^k$

$$\lim_{N\to\infty} \frac{1}{N} Z(B; N) = \mu(B)$$

(here μ stands for the usual k-dimensional Lebesgue measure). Note that it would suffice to consider only boxes $B(b) = B(0, b)$, since the characteristic function of every other box can be obtained by adding and subtracting the characteristic functions of at most 2^k of these special boxes.

Equivalent definitions are given by the following.

Theorem 6.2. *For a sequence (u_n) in $[0, 1)^k$, the following are equivalent*:

(i) *(u_n) is uniformly distributed in $[0, 1)^k$.*

(ii) *For every $f \in R([0, 1]^k)$,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(u_n) = \int_{[0,1]^k} f(x)\,\mathrm{d}x .$$

(iii) (Weyl's criterion) *For every integer point $z \in \mathbb{Z}^k \setminus \{0\}$,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mathrm{e}^{2\pi i z^{\mathrm{T}} u_n} = 0 .$$

Condition (ii) indicates why uniformly distributed sequences are important in the theory of numerical integration. Observe that we obtain an equivalent condition if we assume that (ii) holds for a dense subset of $R([0, 1]^k)$, and Weyl's criterion is obtained by postulating (ii) for the functions $\mathrm{e}^{2\pi i z^{\mathrm{T}} x}$ (and consequently for all linear combinations of these). (This also suggests how the concept of uniformly distribution sequences can be generalized to topological groups.)

As an illustration, we derive from Weyl's criterion the following improvement on the classical Kronecker theorem.

Corollary 6.3 (Weyl 1916). *Suppose that $\alpha_1, \ldots, \alpha_k$ are real numbers such that* 1,

$\alpha_1, \ldots, \alpha_k$ are linearly independent over the rationals. Then the sequence

$$u_n = (\{n\alpha_1\}, \ldots, \{n\alpha_k\}), \quad n \in \mathbb{N}$$

is uniformly distributed in $[0, 1)^k$.

Proof. Let $m = (m_1, \ldots, m_k) \in \mathbb{Z}^k \setminus \{0\}$. Then

$$\sum_{n=1}^{N} e^{2\pi i m^{\mathrm{T}} u_n} = \sum_{n=1}^{N} e^{2\pi i n y},$$

where $y = m_1\alpha_1 + m_1\alpha_2 + \cdots + m_k\alpha_k$. Observe that y is irrational by the hypothesis, and hence $e^{2\pi i y} \neq 1$. Therefore,

$$\left|\sum_{n=1}^{N} e^{2\pi i n y}\right| = \left|e^{2\pi i y} \cdot \frac{1 - e^{2\pi i N y}}{1 - e^{2\pi i y}}\right| \leqslant \frac{2}{|1 - e^{2\pi i y}|} = \mathrm{O}(1).$$

Thus Weyl's criterion is satisfied. □

It is easy to see that the sequence $\omega = (u_n)$, $n \in \mathbb{N}$ is uniformly distributed in $[0, 1)^k$ iff

$$\sup_{\substack{B \subset [0,1)^k \\ \text{aligned box}}} |Z(B, N) - N \cdot \mu(B)| = \mathrm{o}(N).$$

But how small can $\mathrm{o}(N)$ be? To handle this question, put

$$\mathcal{D}_N(B) = Z(N; B) - N|B|,$$
$$\mathcal{D}_N = \sup_{\substack{B \subset [0,1]^k \\ \text{aligned box}}} |\mathcal{D}_N(B)|;$$

and

$$\mathcal{D}_N^p = \left(\int_{[0,1]^k} |\mathcal{D}_N(B(0, x))|^p \, \mathrm{d}x\right)^{1/p}.$$

(Warning: this is *not* the pth power of $\mathcal{D}_N$.)

$\mathcal{D}_N$ and $\mathcal{D}_N^p$ measure (in different norms) the discrepancy of the sequence $u_1, \ldots, u_N$, and their behavior for $N \to \infty$ measures the irregularity of the distribution of the infinite sequence (u_N). In the quantitative theory of uniform distribution, a central problem is the investigation of the order of magnitude of the discrepancy functions $\mathcal{D}_N^p$ and $\mathcal{D}_N$.

The quantitative theory started with the conjecture of Van der Corput (1935a), asserting that for an arbitrary sequence in $[0, 1)$, $\sup_N \mathcal{D}_N = \infty$. This was proved by Van Aardenne-Ehrenfest (1945) who showed that for an arbitrary sequence (u_n)

for infinitely many N,

$$\mathscr{D}_N > c(\log\log N)(\log\log\log N)^{-1}\,.$$

Roth (1954) strengthened this result.

Theorem 6.4. (i) *For an arbitrary infinite sequence (u_n) in $[0,1]^k$ and for every $N > N_0$,*

$$\max_{1\leq n\leq N} \mathscr{D}_n^2 > c_k(\log N)^{k/2}\,.$$

(ii) *For N arbitrary points $u_1, \ldots, u_N$ in $[0,1]^k$,*

$$\mathscr{D}_N^2 > c_k' \,(\log N)^{(k-1)/2}\,.$$

(Here c_k, c_k' are positive constants depending only on k.)

For $k=2$ (Davenport 1956) and for $k\geq 3$ (Roth 1979, 1980) it is proved that (apart from a multiplicative constant) these results are sharp.

The problem of finding bounds for the discrepancy in the supremum norm is more difficult. Since $\mathscr{D}_N \geq \mathscr{D}_N^2$, the preceding results give some lower bounds on $\mathscr{D}_N$. For infinite sequences sharp results are known only for $k=1$, for finite sequences for $k=2$; the latter is a reformulation of Theorem 1.1.

Theorem 6.5 (Schmidt 1972). (i) *For an arbitrary infinite sequence (u_n) in $(0,1)$ and for every $N\geq 2$,*

$$\max_{1\leq n\leq N} \mathscr{D}_N > c\log N\,.$$

(ii) *For arbitrary N points $U = \{u_1, \ldots, u_N\} \subseteq [0,1)^2$,*

$$\mathscr{D}_N > c''\log N\,.$$

(Here c, c' are positive absolute constants.)

This result is best possible apart from the multiplicative constant. If $u_n = \{n\alpha\}$ where α is an irrational number of bounded partial quotients ($a_k \leq K$, $k = 1, 2, \ldots$), then for every N, $\mathscr{D}_N < c_K \log N$. Similarly, for the N points $u_n = \{\{n\alpha\}, n/N\}$ $(1\leq n\leq N)$ in $[0,1]^2$, $\mathscr{D}_N < c_K \log N$.

There is a "transference principle" between *sequences* in $[0,1)^k$ and *sets* in $[0,1)^{k+1}$ (showing that parts (i) and (ii) in both Theorems 6.4 and 6.5 are equivalent). This is given by the following construction.

(1) For a finite sequence $u_1, \ldots, u_N$ in $[0,1)^k$, take the set

$$\left\{\left(u_n, \frac{n-1}{N}\right) \in [0,1)^{k+1} : 1\leq n\leq N\right\}.$$

(2) Let $v_n \in [0,1)^{k+1}$, $1\leq n\leq N$ be N points. Write $v_n = (u_n, y_n)$ where $u_n \in [0,1)^k$ and $y_n \in [0,1)$. Arrange the last coordinates y_n, $1\leq n\leq N$ in increasing order $y_{i_1} \leq y_{i_2} \leq \cdots \leq y_{i_N}$. Take the sequence $u_{i_1}, \ldots, u_{i_N}$ in $[0,1)^k$.

In both cases the discrepancies are the same up to a universal constant factor.

All known proofs of the fundamental Theorem 6.5 are rather hard. We sketch here a proof due to Halász (1981).

Proof of Theorem 6.5. We prove (ii). Given any $x=(x_1,x_2)\in[0,1]^2$, let

$$\mathscr{Z}(x)=|U\cap B(x)|\,,$$

and

$$\mathscr{D}(x)=\mathscr{Z}(x)-Nx_1x_2\,.$$

We shall construct an auxiliary function $F(x)$ such that

$$\left|\int_{[0,1]^2}F(x)\mathscr{D}(x)\,\mathrm{d}x\right|>c_1\log N\,, \tag{6.6}$$

and

$$\int_{[0,1]^2}|F(x)|\,\mathrm{d}x\leq 2\,. \tag{6.7}$$

These yield

$$\mathscr{D}_N\geq\max_x|\mathscr{D}(x)|\geq\tfrac{1}{2}c_1\log n\,,$$

and Theorem 6.5 follows.

Any $x\in[0,1]$ can be written uniquely in the binary form

$$x=\sum_{j=0}^{\infty}\beta_j(x)2^{-j-1}\,,$$

where $\beta_j(x)=0$ or 1 and the sequence $\beta_j(x)$ does not end with $1,1,1,\ldots.$ For $m=0,1,2,\ldots$ let

$$R_m(x)=(-1)^{\beta_m(x)}$$

(Rademacher function). Let $m=(m_1,m_2)$ be a pair of nonnegative integers. Let $\|m\|=m_1+m_2$ and writing $x=(x_1,x_2)$, let

$$R_m(x)=R_{m_1}(x_1)\cdot R_{m_2}(x_2)\,.$$

By an m-box we mean a set of the form

$$[n_1\cdot 2^{-m_1},(n_1+1)\cdot 2^{-m_1}]\times[n_2\cdot 2^{-m_2},(n_2+1)2^{-m_2}]\,.$$

For any m-box A let

$$f_m(x)=\begin{cases}R_m(x)\,, & \text{if } A\cap U=\emptyset\,,\\ 0\,, & \text{otherwise}\,.\end{cases}$$

Let $2N \leqslant 2^n < 4N$, n integer. Let $\alpha = 2^{-6}$, and write

$$F(x) = \prod_{m:\ \|m\| = n} (1 + \alpha f_m(x)) - 1 .$$

Using the orthogonality of the modified Rademacher functions $f_m(x)$, we have

$$\int_{[0,1)^2} |F(x)|\,\mathrm{d}x \leqslant \int_{[0,1]^2} \Big(\prod_{\|m\| = n} (1 + \alpha f_m(x)) + 1 \Big)\,\mathrm{d}x$$

$$= \int_{[0,1)^2} \prod_{\|m\| = n} (1 + \alpha f_m(x))\,\mathrm{d}x + 1 = 1 + 1 = 2 .$$

Note that

$$F(x) = \alpha F_1(x) + \sum_{j=2}^{n+1} \alpha^j F_j(x) ,$$

where

$$F_1(x) = \sum_{\|m\| = n} f_m(x) ,$$

and for $j = 2, \ldots, n+1$

$$F_j(x) = \sum_{\substack{\|m_1\| = \cdots = \|m_j\| = n \\ m_k \neq m_l \text{ if } k \neq l}} f_{m_1}(x) \cdots f_{m_j}(x) .$$

It is not hard to prove that for every m satisfying $\|m\| = n$, we have

$$\int_{[0,1]^2} f_m(x) \mathscr{Z}(x)\,\mathrm{d}x = 0 , \tag{6.8}$$

$$\int_{[0,1]^2} f_m(x) x_1 x_2\,\mathrm{d}x_1\,\mathrm{d}x_2 \geqslant (2^n - N) 2^{-2n-4} , \tag{6.9}$$

and

$$\left| \int_{[0,1]^2} F_j(x) \mathscr{D}(x)\,\mathrm{d}x \right| \leqslant \sum_{k=0}^{n-j+1} \sum_{l=1}^{n-k} 2^{-n-l-4} \cdot N \cdot \binom{l-1}{j-2} . \tag{6.10}$$

The proof of relations (6.8)–(6.10) is straightforward calculation.

Now we are able to complete the proof. By (6.10), we have

$$\left| \sum_{j=2}^{n+1} \alpha^j \int_{[0,1]^2} F_j(x) \mathscr{D}(x)\,\mathrm{d}x \right| \leqslant \sum_{j=2}^{n+1} \sum_{k=0}^{n-j+1} \sum_{l=1}^{n-k} \alpha^j \cdot 2^{-n-l-4} \cdot N \cdot \binom{l-1}{j-2}$$

$$= N\alpha^2 \sum_{k=0}^{n-1} \sum_{l=1}^{n-k} \sum_{j=2}^{n+1} 2^{-n-l-4} \cdot \binom{l-1}{j-2} \alpha^{j-2}$$

$$\leqslant N \cdot \alpha^2 \sum_{k=0}^{n-1} \sum_{l=1}^{\infty} 2^{-n-l-4} (1 + \alpha)^l$$

$$\leqslant N\cdot n\cdot\alpha^2\cdot 2^{-n-4}\sum_{l=1}^{n}\left(\frac{1+\alpha}{2}\right)^l$$
$$\leqslant N\cdot n\cdot\alpha^2\cdot 2^{-n-3}\,.$$

Combining this with (6.8) and (6.9), we obtain

$$\begin{aligned}\left|\int F(x)\mathscr{D}(x)\,\mathrm{d}x\right| &\geqslant \left|\int F_1(x)\mathscr{X}(x)\,\mathrm{d}x_1\,\mathrm{d}x_2\right| - \left|\sum_{j=2}^{n+1}\int_{[0,1]^2} F_j(x)\mathscr{D}(x)\,\mathrm{d}x\right| \\ &\geqslant \alpha(n+1)N(2^n-N)\cdot 2^{-2n-4} - N\cdot n\cdot\alpha^2\cdot 2^{-n-3} \\ &> 2^{-15}\cdot n = c\log N\,,\end{aligned}$$

as required. □

As to the discrepancy in supremum norm, the following is a very difficult old problem.

Conjecture 6.11. For all $k\geqslant 2$ and for N arbitrary points in $[0,1]^k$,

$$\mathscr{D}_N > c(k)(\log N)^{k-1}\,.$$

This would mean that the exponent $(k-1)/2$ implied by Theorem 6.4 (using $\mathscr{D}_N\geqslant\mathscr{D}_N^2$) is only half the truth. Note that the case $k=2$ is settled by Theorem 6.5. If true, Conjecture 6.6 is best possible by the Van der Corput–Halton–Hammersley sequence, see, e.g., Beck and Chen (1987). Recently Beck (1989a) improved on the old result of Roth by proving a 2-dimensional version of the Aardenne-Ehrenfest theorem, but Conjecture 6.11 appears still very difficult.

Approximation of measures. One interpretation of Theorem 6.5 is that it is impossible to approximate the Lebesgue measure on the system of rectangles "too well" with a measure of finite support. There is a more general phenomenon in the background, as proved by Chen: the same is true for arbitrary measures.

Theorem 6.12 (Chen 1984). *Let g be a Lebesgue-integrable function in E^2, and assume that $g(x)\neq 0$ on a subset $S\subseteq E^2$ with $\mu(s)>0$. Then there exists a constant $c(g)>0$ such that for every set U of N points in E^2 and for every function $\lambda: U\to\mathbb{R}$,*

$$\sup_{x\in E^2}\left|\sum_{u\in B(x)\cap U}\lambda(u) - N\int_{B(x)} g(x)\,\mathrm{d}\mu\right| > c(g)\log N\,.$$

Rectangles in the $N\times N$ lattice. It is easy to see that Schmidt's theorem 6.5 has the following corollary. Let the hypergraph $\mathscr{L}_N$ be defined on the underlying set

$S = \{0, 1, \ldots, N\}^2$ by

$$\mathcal{S}_N = \{S \cap B(a, b) \,|\, 0 \leq a \leq N, 0 \leq b \leq N\} .$$

Obviously $\mathcal{D}(\mathcal{S}_N) = 1$. What can be said about $\mathcal{D}_D(\mathcal{S}_N)$, $\mathcal{D}_I(\mathcal{S}_N)$ or $\mathcal{D}_H(\mathcal{S}_N)$? It follows easily from Theorem 6.5 that with a positive constant $c > 0$

$$\mathcal{D}_D(\mathcal{L}_N) > c \log N .$$

Hence by Theorem 3.1,

$$\mathcal{D}_H(\mathcal{L}_N) \geq \tfrac{1}{2}\mathcal{D}_I(\mathcal{L}_N) \geq \tfrac{1}{2}\mathcal{D}_D(\mathcal{L}_N) > c \log n .$$

A related problem concerning balanced 2-colorings of finite sets in the plane was formulated by G. Tusnády. Let $\mathcal{P}$ be an N-element point set on the plane. Let $T = T(\mathcal{P})$ be the least integer t such that one can assign ±1s to the points of P so that the sum of these values in any rectangle with sides parallel to the coordinate axes has absolute value at most T. Now Tusnády's problem is to determine

$$T_N = \max_{|\mathcal{P}| = N} T(\mathcal{P}) .$$

The following theorem gives the best known bounds; the lower bound is due to Beck (1981a), the upper is a recent result of Bohus (1990), improving a result of Beck.

Theorem 6.13. *For* $N \geq 2$,

$$c_1 \cdot \log N < T_N < c_2 \cdot (\log N)^3 .$$

Proof. We give the proof of Beck's upper bound of $(\log N)^4$ as an application of Theorem 2.2. It suffices to prove the following. Let $A = (a_{ij})$, where $a_{ij} = 0$ or 1, be a matrix of size $N \times N$. Then there exist "signs" $\varepsilon_{ij} \in \{-1, +1\}$ such that

$$\left| \sum_{i=1}^{n} \sum_{j=1}^{m} \varepsilon_{ij} a_{ij} \right| < c(\log N)^4 \tag{6.14}$$

for all $1 \leq n$, $m \leq N$.

We now prove (6.14). Adding a few 0-rows and 0-columns if necessary, we may assume that $N = 2^l$ where l is an integer. For every pair (p, q) of integers satisfying $0 \leq p$, $q \leq l$, we partition A into 2^{p+q} submatrices, splitting the horizontal side of the matrix into 2^p equal pieces and the vertical side of the matrix into 2^q equal pieces. There are $(l + 1)^2 \approx (\log N)^2$ partitions. Let us call a submatrix of A special if it occurs in one of these partitions. Let $S = \{(i, j): a_{ij} = 1\}$, and let us associate with every submatrix B the subset

$$Y_B = \{(i, j): a_{ij} \text{ belongs to } B, a_{ij} = 1\} .$$

Let

$$\mathcal{H} = \{Y_B\colon B \text{ is a special submatrix of } A\} .$$

Since the maximum degree $\Delta(\mathcal{H}) \leqslant (l+1)^2$, by Theorem 2.2 there exists an assignment of ±1s so that the absolute value of the sum of the signed entries in each of the special submatrices is less than $2\Delta(\mathcal{H}) \leqslant 2(l+1)^2$. Note, however, that any submatrix of A containing the lower corner A is the union of at most l^2 disjoint special submatrices. Thus (6.14) follows.

The proof of the lower bound depends on Theorems 3.1 and 6.5. We may clearly assume that $N = n^2$, n integer. We need the following reformulation of Theorem 6.5:

Let P be an arbitrary finite set in the square $[0, y)^2$, $y > 1$. There exists an aligned rectangle $A \subset [0, y)^2$ such that

$$\big||P \cap A| - \mu(A)\big| > c \cdot \log y .$$

We shall use that for any convex set $A \subset [0, n)^2$, we have

$$|A \cap \mathbb{Z}^2| = \mu(A) + \mathrm{O}(n) .$$

Let $S = [0, n)^2 \cap \mathbb{Z}^2$, $\mathcal{H} = \{S \cap A\colon\ A \subset [0, n)^2 \text{ aligned rectangle}\}$ and $q = 1 - 2 \cdot n^{-1}$. Let $\varepsilon(s) \in \{-1, +1\}$ $(s \in S)$ be fixed such that

$$\mathcal{D}(\mathcal{H}; q) = \max_A \left|\sum_{s \in S \cap A} (\varepsilon(s) - q)\right| ,$$

and let $S^- = \{s \in S\colon\ \varepsilon(s) = -1\}$. Then we have

$$\begin{aligned}\mathcal{D}(\mathcal{H}; q) &= 2 \max_A \left||S^- \cap A| - \frac{1}{n}|A \cap \mathbb{Z}^2|\right| \\ &\geqslant 2 \max_A \left||S^- \cap A| - \frac{1}{n}\mu(A)\right| + \mathrm{O}(1) .\end{aligned}$$

Apply a contraction of linear ratio $n^{-1/2} : 1$, and apply the reformulation of Schmidt's theorem given above to the resulting set. We obtain that

$$\mathcal{D}(\mathcal{H}; q) \geqslant 2c \cdot \log(n^{1/2}) - \mathrm{O}(1) \geqslant c \log N .$$

Thus by Theorem 3.1,

$$\mathcal{D}_{\mathrm{H}}(\mathcal{H}) \geqslant \tfrac{1}{2}\mathcal{H}_1(\mathcal{H}) \geqslant \tfrac{1}{2}\mathcal{D}(\mathcal{H}; q) \geqslant c_2 \cdot \log N .$$

In other words,

$$\mathcal{D}(\mathcal{H}_Z) \geqslant c_2 \cdot \log N$$

for some $Z \subset S$. Since $|Z| \leqslant |S| = n^2 = N$, we have

$$T_N \geqslant T_{|Z|} \geqslant \mathcal{D}(\mathcal{H}_Z) \geqslant c_2 \cdot \log N . \qquad \square$$

Let $X = \{(i, j): a_{ij} = 1\}$, $\mathcal{H}$ be the family of submatrices of A containing the lower left corner of A, $\mathcal{G}$ be the family of special submatrices, $N = 2^l$, $D = (l + 1)^2$, $K = l^2$. Applying Theorem 2.7, we obtain the following modest improvement on (6.14):

$$\left|\sum_{i=1}^{n}\sum_{j=1}^{m} \varepsilon_{ij}a_{ij}\right| < c(\varepsilon)\cdot(\log N)^{(7/2)+\varepsilon}$$

for all $1 \leq n$, $m \leq N$.

In higher dimensions, however, the improvement given by Theorem 2.7 becomes significant. Repeating the proof of 6.13 in higher dimensions, we get the following result:

Let $A = (a_n)$ $(n \in \{1, \ldots, N\} \times \cdots \times \{1, \ldots, N\})$ be a k-dimensional matrix with entries $a_n = 0$ or 1. Then there exist "signs" $\varepsilon_n \in \{-1, +1\}$ such that

$$\left|\sum_{n:\, n\leq m} \varepsilon_n a_n\right| \leq c(k)\cdot(\log N)^{2k}$$

for all $m = (m_1, m_2, \ldots, m_k)$ satisfying $1 \leq m_1, \ldots, m_k \leq N$. Applying Theorem 2.7 we obtain, however, the following better bound:

$$\left|\sum_{n:\, n\leq m} \varepsilon_n a_n\right| < c(k, \varepsilon)\cdot(\log N)^{k+(3/2)+\varepsilon}\,,$$

for all m as above. We have strong indications that the true order of magnitude is probably about $(\log N)^{k-1}$.

In contrast to the case of Theorem 2.2, when the proof gives a polynomially computable algorithm to construct the desired signs $\varepsilon_i = \pm 1$, the proofs of Theorems 2.5 and 2.7 imply only the *existence* of balanced 2-colorings.

7. Geometric structures

In this section we discuss a variety of questions where the underlying set S is either the k-dimensional unit cube $[0, 1]^k$ or (in the discrete version) the $N \times N \times \cdots \times N$ lattice.

We study generalizations of the classical problem considered in Theorem 1.1. We no longer restrict ourselves to the boxes: we allow rotation, and we also study more general shapes. Many problems of this type originated with the paper of Erdős (1964).

Let $\mathcal{A}$ be a family of simple geometric objects, as aligned or tilted rectangles, triangles, balls, etc., in R^k. Let $P \subseteq [0, 1]^k$ be a set of N points. Set

$$\mathcal{D}_N(\mathcal{A}) = \inf_{|P|=N}\ \sup_{A\in\mathcal{A}} \left||A \cap P| - N\mu(A \cap [0, 1]^k)\right|\,.$$

We consider first the case of aligned right triangles.

Theorem 7.1 (Schmidt 1969). *Let $P_1, \ldots, P_N$ be N points in the unit square $[0,1]^2$. Then there exists a right triangle $T \subset [0,1)^2$ with two sides parallel to the coordinate axes, and with*

$$||P \cap T| - N \cdot \mu(T)| > N^{(1/4)-\varepsilon}.$$

Beck (1984a, 1987a) slightly improved the lower bound and also proved that the lower bound is nearly sharp.

Theorem 7.1′. *Let $\mathcal{A}$ be the family of right triangles in the plane with two sides parallel to coordinate axes. Then*

$$c_1 N^{1/4} < \mathcal{D}_N(\mathcal{A}) < N^{1/4}\sqrt{\log N}.$$

This theorem exhibits a rather paradoxical phenomenon. Let T' be a right triangle. There is a unique right triangle T'' such that $T' \cup T''$ is an aligned rectangle A. We know that there exist N-element sets P with

$$||P \cap A| - N \cdot \mu(A)| < c \cdot \log N$$

for all aligned rectangles $A \subset [0,1)^2$. This set contains almost the "right" number of points in $T' \cup T''$ but – by Theorem 7.1 – must be quite irregularly distributed in the two halves T' and T''.

Essentially the same proof gives the following 2-coloring result. Let f be a 2-coloring of the $N \times N$ square lattice. Then there exists an aligned right triangle T such that the difference between the number of red points and the number of blue points in T is at least $c \cdot N^{1/2}$. In other words, the corresponding hypergraph has discrepancy at least $c \cdot N^{1/2}$.

(Note that the analogous question for aligned rectangles is trivial. The chessboard type 2-coloring of $N \times N$ has deviation at most 1 for any aligned rectangle.)

Consider next the family of balls. Again we have a "large discrepancy" result (for the pioneering result, see Schmidt 1969).

Theorem 7.2 (Beck 1987a). *Let $\mathcal{A}$ be the family of balls contained in $[0,1)^k$. Then*

$$\mathcal{D}_N(\mathcal{A}) > N^{1/2-1/2k-\varepsilon}.$$

The following result states, roughly speaking, that for rotation invariant families the discrepancy is always "large".

Theorem 7.3 (Beck 1987a). *Let $A \subset [0,1)^k$ be a k-dimensional convex body, and let $\mathcal{A}$ be the family of convex sets obtained from A by a similarity transformation (rotation, translation, and homothetic transformation). Then*

$$\mathcal{D}_N(\mathcal{A}) > c(A, \varepsilon) \cdot N^{1/2-1/2k-\varepsilon}.$$

We remark that Theorems 7.2–7.3 are nearly best possible (see Beck 1984a). The situation is more complicated if rotation is forbidden (as the difference between aligned right triangles and rectangles indicates). The discrepancy of the family of homothetic copies of a given convex shape A depends mainly on the smoothness of the boundary A. We have a fairly good understanding of this phenomenon (for more details, see Beck 1988 and Beck and Chen 1987).

For the discrepancy of congruent sets, see Beck (1987b). For the discrepancy of half-plances, see Beck (1983), Alexander (1990) and Matousek (1994).

It is worthwhile to mention here that all of these theorems are essentially independent of the shape of the underlying set – instead of the unit cube one can consider the unit ball, the regular simplex, any "reasonable" convex body, the surface of the unit sphere, etc.

An application in discrete geometry. For which set of N points on the unit sphere is the sum of all $\binom{N}{2}$ euclidean distances between these points maximal, and what is the maximum? Let S^k denote the surface of the unit sphere in $\mathbb{R}^{k+1}$. Let P be a set of N points on S^k. Let $|x|$ denote the usual euclidean length. We define

$$L(N, k, P) = \sum_{p,q \in P} |p - q|$$

and

$$L(N, K) = \max_{P} L(N, k, P) ,$$

where the maximum is taken over all $P \subset S^k$, $|P| = N$. The determination of $L(N, k)$ is a long-standing open problem in discrete geometry. For $k = 1$, the solution is given by the regular N-gon. It is also known that for $N = k + 2$, the regular simplex is optimal. For $N > k + 2$ and $k \geqslant 2$, the exact value of $L(N, k)$ is unknown. The reason for this is that if N is sufficiently large compared to k, then there are no "regular" configurations on the sphere, so the extremal point system(s) is (are), as expected, quite complicated and "ad hoc".

Since the determination of $L(N, k)$ seems to be hopeless, it is natural to compare the discrete sum $L(N, k, P)$ with the following integral (the solution of the "continuous relaxation" of the distance problem)

$$\frac{N^2}{2} \cdot \frac{1}{\sigma(S^k)} \int_{S^k} |p - p_0| \, \mathrm{d}\sigma(P) = c_0(k) \cdot N^2 ,$$

where σ denotes the surface area, $\mathrm{d}\sigma(P)$ represents an element of the surface area on S^k, $p_0 = (1, 0, 0, \ldots, 0) \in \mathbb{R}^{k+1}$. The constants $c_0(k)$ can be calculated explicitly; e.g., $c_0(1) = 2/\pi$, $c_0(2) = \frac{2}{3}$). Stolarsky (1973) has discovered a beautiful identity saying, roughly speaking, that the discrete sum $L(N, k, P)$, plus a measure of how far the set P deviates from uniform distribution, is constant. Thus the sum of distance is maximized by a well-distributed set of points. Combining Stolarsky's identity with a result in "irregularities of distribution", one can obtain some nontrivial information on the order of magnitude of $L(N, k)$ (see Beck 1984b).

Theorem 7.4. $L(N, k) = c_0(k) \cdot N^2 + O(N^{1-(1/k)})$.

Finally, we mention the famous Heilbronn's triangle problem which is, in a broader sense, related to our topic (see Roth 1976 and Komlós et al. 1982).

8. Uniform distribution and ergodic theory

The most important class of uniformly distribution sequences in $[0, 1)$ is the class of sequences $(\{n\alpha\})$ for α irrational. These are the basic sequences in the theory of diophantine approximation. Further, these are the best "test-sequences": very often theorems which were found first for sequences $(\{n\alpha\})$ turned out to be true for more general ones. Finally we mention the relation of sequences $(\{n\alpha\})$ to topological transformations.

The discrepancy of $((n\alpha\})$ depends on the partial quotients a_k, $k = 1, 2, \ldots$ of α. For every N and $x \in [0, 1)$ there is an "explicit" formula for the discrepancy $\mathcal{D}_N([0, x))$ defined in section 6 (Sós 1974). This leads to the following bounds.

Theorem 8.1. *Let p_k/q_k be the kth convergent of α: $p_k/q_{jk} = [a_1, \ldots, a_{k-1}]$. If $q_k \leq N < q_{k+1}$ then*

$$c_1 \sum_{i=1}^{k} a_i < \max_{1 \leq n \leq N} \mathcal{D}_N < c_2 \sum_{i=1}^{k+1} a_i .$$

Consequently, if $a_i \leq K$, $i = 1, \ldots$, then

$$\mathcal{D}_N < c \cdot K \cdot \log N .$$

Much is known about the finer properties of the distribution. Though

$$\max_{1 \leq n \leq N} \sup_{I} \mathcal{D}_n(I) > c \log N ,$$

there are intervals I in which the distribution is very good.

Theorem 8.2 (Hecke–Kesten). *For the sequence $(\{n\alpha\})$ and for a fixed interval I, the discrepancy $\mathcal{D}_N(I)$ remains bounded if and only if $\mu(I) = \{k\alpha\}$ for some integer k.*

The "if" part was proved by Hecke (1922) and the much deeper "only if" part by Kesten (1966). Very elegant proofs and generalizations of this theorem in the framework of ergodic theory are due to Fürstenberg et al. (1973), Halász (1976), Petersen (1973).

On the other hand it is remarkable that this theorem and further properties of $\mathcal{D}_N$ are relevant in ergodic theory (see, e.g., Herman 1976, Deligne 1975).

Schmidt investigated the analogous question for arbitrary sequences in $[0, 1)$.

Theorem 8.3 (Schmidt 1974). *For an arbitrary sequence (u_n) in $[0,1)$ the lengths of all intervals I for which $\mathscr{D}_N(I)$ remains bounded form a countable set.*

The ergodic theoretical generalization shows the essence of Kesten's theorem.

Let $(\Omega, \mathscr{A}, \mu)$ be a probability space, $T:\Omega \to \Omega$ an ergodic transformation (a measure preserving transformation such that every T-invariant measurable set has measure 0 or 1). For $A \in \mathscr{A}$, $x \in \mathscr{A}$ let $Z_N^T(A;x)$ denote the number of points $T^n x \in A$, $1 \leqslant n \leqslant N$. Set

$$\mathscr{D}_N^T(A;x) = |Z_N^T(A;x) - N\mu(A)|\,.$$

By Birkhoff's ergodic theorem, for every fixed $A \in \mathscr{A}$, for almost all $x \in \Omega$,

$$\frac{1}{N}\mathscr{D}_N^T(A;x) \to 0 \quad \text{if } N \to \infty\,.$$

The uniformity or irregularity of the distribution of the orbit is measured by the sequence $\mathscr{D}_N^T(A;x)$. Fürstenberg et al. (1973), Petersen (1973), Halász (1976) proved the following very striking generalization of Kesten's theorem.

Theorem 8.4. *If for some $A \in \mathscr{A}$, $\mathscr{D}_N^T(A;x)$ is bounded on a set $X \subset \Omega$ of positive measure, then $\mathrm{e}^{2\pi \mathrm{i}\mu(A)}$ is an eigenvalue of T; that is, there exists a function $g \not\equiv 0$ such that*

$$g(Tx) = \mathrm{e}^{2\pi \mathrm{i}\mu(A)} g(x) \quad \textit{for } x \in \Omega\,.$$

Conversely, for every eigenvalue $e^{2\pi i\mu}$ there exists an $A \in \mathscr{A}$ such that $\mu(A) = \mu$ and $\mathscr{D}_N^T(A;x)$ remains bounded as $N \to \infty$ for almost all $x \in \Omega$.

Remark. Kesten's theorem follows from Theorem 8.4. To see this, let $\Omega = \mathbb{R}/\mathbb{Z}$. Let $R_\alpha : x \to x + \alpha \pmod 1$ be the rotation by $2\pi\alpha$). The eigenvalues of R_α are the numbers $\mathrm{e}^{2\pi \mathrm{i}\{k\alpha\}}$; hence Kesten's theorem follows.

We give another example of the relationship between uniform distribution and ergodic theory, illustrating how results on distribution of the sequences $(\{n\alpha\})$ imply general results on homeomorphisms of the circle.

Denjoy (1932) proved that for every homeomorphism $T: \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ having no periodic point there exists an irrational $\alpha(T) \in (0,1)$ such that T is conjugate to the rotation R_α. By this result, the distribution of $T^n x$, $n = 1, 2,, \ldots,$ is determined by the distribution of the sequence $(\{n\alpha\})$. In particular,

(a) By Birkhoff's ergodic theorem the discrepancy $\mathscr{D}_N^T(I;x) = \mathrm{o}(N)$. By Denjoy's theorem we know much more: $\mathscr{D}_N^T(I;x)$ is the same as the corresponding discrepancy of the sequence $(\{n\alpha(T)\})$.

(b) The order of points $\{n\alpha\}$, $1 \leqslant n \leqslant N$ is very restricted: if π is the permutation determined by $\{\pi(1)\alpha\} < \cdots < \{\pi(N)\alpha\}$, then, for example, for every fixed α and N the difference $\pi(i) - \pi(i-1)$ takes at most three different values. Now, by Denjoy's theorem the same holds for an arbitrary homeomorphism T having no periodic point and every point x, if we define the permutation π by $T^{\pi(1)}(x) < T^{\pi(2)}(x) < \cdots < T^{\pi(N)}$. (See Sós 1957, Swierczkowski 1958.)

One of the most fascinating and deepest relationships between combinatorics and ergodic theory is given by Fürstenberg. Since there is a recent expository paper by Fürstenberg et al. (1982), and the book of Fürstenberg (1981), we do not go into the discussion of this. We mention only the fascinating recent result of Fürstenberg and Katznelson (1989) on the density version of the Hales–Jewitt theorem (see chapter 25).

9. More versions of discrepancy

Strong irregularity. In $[0, 1)$ the following "strong irregularity" phenomenon holds.

Theorem 9.1. (i) *For every* $\varepsilon > 0$ *there exists a* $\delta > 0$ (*depending only on* ε) *such that for an arbitrary sequence* (u_n) *in* $(0, 1)$ *and for every* $N > 0$, $\mathcal{D}_n > \delta \log n$ *for all but at most* N^{ε} *values of* $n \leq N$.

(ii) *For every* $K > 0$ *there exists an* $M > 0$ (*depending only on* K) *such that for an arbitrary sequence* (u_n) *in* $(0, 1)$ *and for every* $N > 0$, $\mathcal{D}_n > K$ *for all but at most* $(\log N)^M$ *values of* $n \leq N$.

(iii) *For an arbitrary sequence* (u_n) *in* $(0, 1)$ *the set of values of* x *for which* $\mathcal{D}_N([0, x)) = o(\log N)$ *holds, has Hausdorff dimension* 0.

This theorem was proved first only for $(\{n\alpha\})$ sequences (Sós 1979, 1983a), then for arbitrary sequences and in a more general form by Halász (1981) and Tijdeman and Wagner (1980).

One-sided irregularities. Measuring the irregularities with $\mathcal{D}_N$ or $\mathcal{D}_N^p$, we do not have any information on the sign of the discrepancy. Therefore we define

$$\mathcal{D}_N^+([0, \beta)) = \max\{\mathcal{D}_N([0, \beta)), 0\} ,$$

and

$$\mathcal{D}_N^+ = \sup_{\beta} \mathcal{D}_N^+([0, \beta)) .$$

$\mathcal{D}_N^-$ is defined analogously.

One-sided discrepancies show some new phenomena. Again, the first results on one-sided irregularities were found for $(\{n\alpha\})$ sequences. For example, there is *no* one-sided strong irregularity phenomenon. We mention just the simplest illustration of this. It is easy to see that

$$\sup_N \mathcal{D}_N^+ = \infty , \qquad \sup_N \mathcal{D}_N^- = \infty .$$

But no explicit lower bound can be given: for an arbitrary sequence $M_N \to \infty$ there exists an α such that $\mathcal{D}_N^+ < M_N$, and also there exists an α such that $\mathcal{D}_N^- < M_N$, if N is large enough.

Similarly, it is easy to see that the sequence of indices N with $\mathcal{D}_N^+ < K$ has density 0. However, for an arbitrary sequence $M_N = \mathrm{o}(N)$, there exist an α and a K such that $\mathcal{D}_N^+ < K$ holds for at most M_N values of $n < N$, if N is large enough (Sós 1983a).

Concerning intervals of small discrepancy, first we remark that $\mathcal{D}_N^+([0, \beta))$ may be bounded even in the case when $\beta \neq \{k\alpha\}$, i.e. when $\mathcal{D}_N([0, \beta))$ is not.

In Dupain and Sós (1978) those intervals $[0, \beta)$ are investigated for which $\mathcal{D}_N^+([0, \beta))$ is bounded. Here we mention just one of the new phenomena: there exists an α for which the set $\{\beta\colon \sup_N \mathcal{D}_N^+([0, \beta)) < \infty\}$ has the cardinality continuum.

As an example in the opposite direction, the assertion in Theorem 8.2 remains true if instead of the boundedness of $\mathcal{D}_N(A)$ we assume only one-sided boundedness. Halász (1976) proved that if

$$\sup_N \mathcal{D}_N^+(A; x) < \infty$$

holds on a set $X \subset \Omega$ of positive measure, then $\mathrm{e}^{2\pi\mathrm{i}\mu(A)}$ must be an eigenvalue of T.

In contrast to aligned boxes, for balls even the simplest results: $\sup_N \mathcal{D}_N^+ = \infty$, $\sup_N \mathcal{D}_N^- = \infty$ are nontrivial. The proof of these, that is, a one-sided version of Theorem 7.2, can be found in Beck (1989b).

The following problem of Erdős, which was recently solved, is essentially a one-sided discrepancy problem.

Let $\xi_1, \xi_2, \xi_3, \ldots$ be an arbitrary infinite sequence of complex numbers on the unit circle $|z| = 1$. For every $n \in \mathbb{N}$ and complex z, let

$$P_n(z) = \prod_{j=1}^{n} (z - \xi_j) .$$

Further, let

$$A_n = A(\xi_1, \xi_2, \ldots, \xi_n) = \max_{|z|=1} |P_n(z)| .$$

Erdős conjectured that for every fixed sequence (ξ_n), $\limsup A_n = \infty$, and asked about the correct order of magnitude of

$$\max_{1 \leq n \leq N} A_n \quad \text{as } N \to \infty .$$

Observe that if the points $\xi_1, \ldots, \xi_n$ are just the nth roots of unity, then $P_n(z) = z^n - 1$, and so $A_n = 2$. This shows that the relation $\limsup A_n = \infty$ must be a consequence of the impossibility of getting every segment $\xi_1, \ldots, \xi_n$ close to uniform distribution. There seems, therefore, to be an intimate connection with the Van der Corput problem (see section 6).

By realizing this heuristics, Wagner (1980) proved the conjecture $\limsup A_n =$

∞. He developed a variation of Schmidt's original proof of Theorem 6.5, and actually proved the estimate

$$\max_{1\leq n\leq N} A_n > (\log N)^c .$$

Recently, Beck (1991a) managed to prove the best possible result

$$\max_{1\leq n\leq N} A_n > N^c ,$$

by developing a version of Halász's proof of Theorem 6.5.

10. Epilogue

As we mentioned already in the introduction, discrepancy theory has its roots, as well as its applications, in many different areas. Here we mention just a few recent applications of discrepancy and uniform distribution.

Squaring the circle. Tarski raised the following question, which is sometimes called "the problem of squaring the circle" (misusing the name of an ancient problem): is a disc equidecomposable to a square? In other words, can a disc be decomposed into finitely many parts, which can be arranged to obtain a partition of a square? The answer is in the negative under various restrictions, e.g., if the pieces are restricted to be Jordan domains.

Recently Laczkovich (1990) gave a striking and ingenious construction which answers Tarski's question in the affirmative. The proof is based on a sufficient condition for the equidecomposability of two bounded measurable sets in terms of the discrepancy of certain special sequences.

Computing the volume. Uniformly distributed sequences are used generally in applications of Monte Carlo methods. A recent success in this area is the computation of the volume of an n-dimensional convex body in polynomial time by Dyer et al. (1989). The basic tool is that a uniformly distributed point in the body can be generated efficiently (using random walk on a grid). It is a surprising fact that in this problem deterministic uniformly distributed sequences cannot give a good approximation in polynomial time (see Elekes 1986, Bárány and Füredi 1987).

Drawing segments on screen. Luby (1986) studied the question of drawing segments on a screen as paths in a grid. He showed that if certain natural assumptions are made, every scheme to assign a "connecting segment" to every pair of points will necessarily use "bent" segments. The amount of deviation from the straight line is determined by Schmidt's theorem 1.1.

"Gray" areas in photography. Rödl and Winkler (1990) studied the question of representing a gray area as a combination of black and white dots. Modelling the

"smoothness" of the resulting color as a discrepancy problem, he showed that the measure of this "smoothness" can be estimated by the theorem of Beck and its improvement by Bohus (Theorem 6.13).

References

Alexander, R.
[1990] Geometric methods in the study of irregularities of distribution, *Combinatorica* **10**, 115–136.

Alon, N., D.J. Kleitman, C. Pomerance, M. Saks and P. Seymour
[1987] The smallest n-uniform hypergraph with positive discrepancy, *Combinatorica* **7**, 151–160.

Bárány, I.
[1981] A vector-sum theorem and its application to improving flow shop guarantees, *Math. Oper Res.* 6(3), 445–452.

Bárány, I., and Z. Füredi
[1987] Computing the volume is difficult, *Discrete Comput. Geom.* **2**, 319–326.

Bárány, I., and V.S. Grinberg
[1981] On some combinatorial questions in finite dimensional spaces, *Linear Algebra Appl.* **41**, 1–9.
[1985] A vector-sum theorem in two-dimensional space, *Period. Math. Hung.* **16**(2), 135–138.

Baranyai, Zs.
[1974] On the factorization of the complete uniform hypergraph, in: *Infinite and Finite Sets*, eds. A. Hajnal, R. Rado and V.T. Sós, *Colloq. Math. Soc. János Bolyai* **10**, pp. 91–108.

Beck, J.
[1981a] Balanced two-colorings of finite sets in the square, *Combinatorica* 1(4), 327–335.
[1981b] Roth's estimate of the discrepancy of integer sequences is nearly sharp, *Combinatorica* 1(4), 319–325.
[1983] On a problem of K.F. Roth concerning irregularities of point distributions, *Invent. Math.* **74**, 477–487.
[1984a] Some upper bounds in the theory of irregularities of distribution, *Acta Arithm.* **43**, 115–130.
[1984b] Sums of distances between points on a sphere an application of the theory of irregularities of distribution to discrete geometry, *Mathematika* **31**, 33–41.
[1987a] Irregularities of Distribution I, *Acta Math.* **59**, 1–49.
[1987b] On a problem of Erdős in the theory of irregularities of distributions, *Math. Ann.* **277**, 233–247.
[1988] Irregularities of Distribution II, *Proc. London Math. Soc.* **56**(3), 1–50.
[1989a] A 2-dimensional van Aardenne-Ehrenfest theorem in irregularities of distribution, *Compos. Math.* **72**, 269–339.
[1989b] On a problem of W.M. Schmidt concerning one-sided irregularities of point distributions, *Math. Ann.* **285**, 29–55.
[1990] *On Sums of Infinite-Dimensional Vectors*, Manuscript, to appear in David Reimer's Ph.D. Thesis (Rutgers University, Piscataway, NJ).
[1991a] The modulus of polynomials with zeros on the unit circle – a problem of Erdős, *Ann. of Math.* **134**, 609–651.
[1991b] Flat polynomials on the unit circle – on a problem of Littlewood, *Bull. London Math. Soc.* **23**, 269–277.

Beck, J., and W. Chen
[1987] *Irregularities of Distribution* (Cambridge University Press, Cambridge).

Beck, J., and T. Fiala
[1981] "Integer-making" theorems, *Discrete Appl. Math.* **3**, 1–8.

Beck, J., and J. Spencer
[1983] Balancing matrices with line shifts, *Combinatorica* 3(3–4), 299–304.
[1984a] Well-distributed 2-colorings of integers relative to long arithmetic progressions, *Acta Arithm.* **43**, 287–294.
[1984b] Integral approximation sequences, *Math. Programming* **30**, 88–98.

[1989] Balancing matrices with line shifts, II, in: *Irregularities of Partitions, Algorithms and Combinatorics*, Vol. 8 (Springer, Berlin) pp. 23–37.

Bohus, G.

[1990] On the discrepancy of 3 permutations, *Random Structures and Algorithms* **1**, 215–220.

Davenport, H.

[1956] Note on irregularities of distribution, *Mathematika* **3**, 131–135.

Deligne, P.

[1975] *Les Difféomorphismes du Cercle, Séminaire Bourbaki, Lecture Notes in Mathematics*, Vol. 477, 01 (Springer, Berlin).

Denjoy, A.

[1932] Sur les courbes définies par des équations differentielles à la surface du tore, *J. Math. Pures Appl.* **11**, 333–375.

Dupain, Y., and V.T. Sós

[1978] On the one-sided boundedness of discrepancy function of the sequence $(n\alpha)$, *Acta Arithm.* **37**, 363–374.

Dvoretzky, A.

[1961] Some results on convex bodies and Banach spaces, in: *Proc. Symp. on Linear Spaces, Jerusalem* (Pergamon Press, Oxford) pp. 123–160.

Dyer, M., A. Frieze and R. Kannan

[1989] A random polynomial time algorithm for approximating the volume of convex bodies, in: *Proc. 21st Annu. ACM Symp. on Theory of Computing* (ACM, New York) pp. 375–381.

Elekes, G.

[1986] A geometric inequality and the complexity of computing volume, *Discrete Comput. Geom.* **1**, 289–292.

Erdős, P.

[1964] Problems and results on diophantine approximations, *Comput. Math.* **16**, 52–65.

Erdős, P., and J. Spencer

[1972] Inbalances in k-colorations, *Network* **1**, 379–385.

Fürstenberg, H.

[1981] *Recurrence in Ergodic Theory and Combinatorial Number Theory* (Princeton University Press, Princeton, NJ).

Fürstenberg, H., and Y. Katznelson

[1989] Idempotents in compact semigroups and Ramsey theory, *Israel J. Math.* **68**, 257–270.

Fürstenberg, H., H. Keynes and L. Shapiro

[1973] Prime flows in topological dynamics, *Israel J. Math.* **14**(1), 26–38.

Fürstenberg, H., Y. Katznelson and D. Ornstein

[1982] The ergodic theoretical proof of Szemerédi's theorem, *Bull. Amer. Math. Soc.* 7(3), 447–654.

Ghouila-Houri, A.

[1962] Caractérisation des matrices totalement unimodulaires, *C.R. Acad. Sci. Paris* **254**, 1192–1194.

Hajela, D.

[1988] On a conjecture of Komlós about signed sums of vectors inside the sphere, *European J. Combin.* **9**, 33–37.

Halász, G.

[1976] Remarks on the remainder in Birkhoff's ergodic theorem, *Acta Math. Hungar.* **27**, 389–396.

[1981] On Roth's method in the theory of irregularities of point-distributions, in: *Proc. Conf. on Analytic Number Theory, Durham*, Vol. 2 (Academic Press, London) pp. 79–94.

Hecke, E.

[1922] Über analytische Funktionen und die Verteilung von Zahlen mod Eins, *Abh. Math. Sem. Hamburg* **1**, 54–76.

Herman, M.R.

[1976] Conjugaison c^{∞} des difféomorphismes du cercle dont le nombre de rotation satisfait a une condition arithmétique, *C.R. Acad. Sci. Paris* **282**.

Hoffman, A.

[1987] Oral communication.

Kesten, H.

[1966] On a conjecture of Erdős and Szüsz related to uniform distribution mod 1, *Acta Arithm.* **12**, 193–212.

Komlós, J., and M. Sulyok

[1970] On the sum of elements of ±1-matrices, in: *Combinatorial Theory and Its Applications, Proc. Colloq. Balatonfüred, Hungary* (North-Holland, Amsterdam) pp. 721–728.

Komlós, J., J. Pintz and E. Szemerédi

[1982] A lower bound for Heilbronn's problem, *J. London Math. Soc. (2)* **25**, 13–24.

Kuipers, L., and H. Niederreiter

[1974] *Uniform Distribution of Sequences, Pure and Applied Mathematics* (Wiley Interscience, New York).

Laczkovich, M.

[1990] Equidecomposability and discrepancy; a solution of Tarski's circle squaring problem, *J. Reine Angew. Math.* **404**, 77–117.

Lovász, L., J. Spencer and K. Vesztergombi

[1986] Discrepancy of set-systems and matrices, *European J. Combin.* 7, 151–160.

Luby, M.

[1986] *Grid Geometries which Preserve Properties of Euclidean Geometry: A Study of Line Drawing Algorithms,* Technical Report 190/86 (University of Toronto, Computer Science Department).

Matousek, J.

[1994] *Tight upper Bounds for the Discrepancy of Half-planes,* Manuscript.

Matousek, J., and J. Spencer

[1994] *Discrepancy in Arithmetic Progressions,* Manuscript.

Olson, J.E., and J. Spencer

[1978] Balancing families of sets, *J. Combin. Theory A* **25**, 29–37.

Petersen, K.

[1973] On a series of cosecants to a problem in ergodic theory, *Comp. Math.* **26**(3), 313–317.

Ramsey, F.P.

[1930] On a problem of formal logic, *Proc. London Math. Soc. (2)* **30**, 264–285.

Rödl, V., and P. Winkler

[1990] Concerning a matrix approximation problem, *Crux Mathematicorum,* March, pp. 76–79.

Roth, K.F.

[1954] On irregularities of distributions, *Mathematika* **1**, 73–79.

[1964] Remark concerning integer sequences, *Acta Arithm.* **9**, 257–260.

[1976] Developments in Heilbronn's triangle problem, *Adv. in Math.* **22**, 364–385.

[1979] On irregularities of distribution III, *Acta Arithm.* **35**, 373–384.

[1980] On irregularities of distribution IV, *Acta Arithm.* **37**, 67–75.

Sárközy, A.

[1973] in: *Probabilistic Methods in Combinatorics,* eds. P. Erdős and J. Spencer (Akadémiai Kiadó, Budapest, 1974) section 8.

Schmidt, W.M.

[1969] Irregularities of distribution IV, *Invent. Math.* 7, 55–82.

[1972] Irregularities of distribution VII, *Acta Arithm.* **21**, 45–50.

[1974] Irregularities of distribution VIII, *Trans. Amer. Math. Soc.* **198**, 1–22.

Sós, V.T.

[1957] On the theory of diophantine approximation I, *Acta Math. Acad. Sci. Hungar.* **8**, 461–472.

[1974] On the discrepancy of the sequence $(n\alpha)$, *Colloq. Math. Soc. János Bolyai* **13**, 359–367.

[1979] On strong irregularities of the distribution of $\{n\alpha\}$ sequences, *Tagungsber. Oberwolfach* **23**, 17–18.

[1983a] Strong irregularities of the distribution of $(n\alpha)$ sequences I, in: *Studies in Pure Mathematics* (Akadémiai Kiadó).

[1983b] Irregularities of partitions, in: *Surveys in Combinatorics 82, 9th British Combinatorial Conference, 1983,* ed. E.K. Lloyd (Cambridge University Press, Cambridge) pp. 201–245.

Spencer, J.

[1985] Six standard deviation suffice, *Trans. Amer. Math. Soc.* **289**, 679–706.

[1986] Balancing vectors in the max norm, *Combinatorica* **6**, 55–65.

Stolarsky, K.B.

[1973] Sums of distances between points on a sphere II, *Proc. Amer. Math. Soc.* **41**, pp. 575–582.

Swierczkowski, S.

[1958] On successive settings of an arc on the circumference of a circle, *Fund. Math.* **46**, 187–189.

Tijdeman, R., and G. Wagner

[1980] A sequence has almost nowhere small discrepancy, *Monatsh. Math.* **90**, 315–329.

van Aardenne-Ehrenfest, T.

[1945] Proof of the impossibility of a just distribution of an infinite sequence of points over an interval, *Proc. Kon. Nederl. Akad. Wetensch.* **48**, 266–271.

[1949] On the impossibility of a just distribution, *Proc. Kon. Nederl. Akad. Wetensch.* **52**, 734–739.

Van der Corput, J.G.

[1935a] Verteilungsfunktionen I, *Proc. Kon. Nederl. Akad. Wetensch.* **38**, 813–821.

[1935b] Verteilungsfunktionen II, *Proc. Kon. Nederl. Akad. Wetensch.* **38**, 1058–1066.

van der Waerden, B.L.

[1927] Beweis einer Baudetschen Vermutung, *Nieuw Arch. Wisk.* **15**, 212–216.

Wagner, G.

[1980] On a problem of Erdős in diophantine approximation, *Bull. London Math. Soc.* **12**, 81–88.

Weyl, H.

[1916] Über die Gleichverteilung von Zahlen mod Eins, *Math. Ann.* **77**, 313–352.

CHAPTER 27

Automorphism Groups, Isomorphism, Reconstruction

László BABAI*

Department of Algebra, Eötvös University, Budapest H-1088, Hungary

and

Department of Computer Science, University of Chicago, Chicago, IL 60637, USA

Contents

*Section 2 was written in collaboration with W. Imrich

HANDBOOK OF COMBINATORICS
Edited by R. Graham, M. Grötschel and L. Lovász

Note. While this chapter contains a substantial amount of material on infinite graphs, its focus is on finite graphs. Therefore all graphs will be *finite*, unless otherwise stated. Exceptions are sections 3.6, 3.7, and 3.11, where graphs are generally infinite, and sections 3.9, 3.10, 5.3, where a main theme is the interplay between finite and infinite.

Surveys. A portion of the material discussed in this chapter is covered in two survey articles on automorphism groups of graphs: Cameron (1983) and Babai and Goodman (1993). Chapter 12 of Lovász (1979a) is a nice introduction to the subject. A beautiful treatment of the basics of higher symmetry is Biggs (1974). Brouwer et al. (1989) is a monumental yet enjoyable work on distance-transitivity and related subjects. Much of our current knowledge on graph isomorphism testing is summarized in Babai and Luks (1983). The general concept of reconstruction (invertibility of various constructions) is illustrated in chapter 15 of Lovász (1979a). Recent surveys on the Kelly–Ulam graph reconstruction conjecture include Bondy (1991), Ellingham (1988) (see also Bondy and Hemminger 1977).

0. Introduction

0.1. Graphs and groups

A study of graphs as geometric objects necessarily involves the study of their symmetries, described by the group of automorphisms. Indeed, there has been significant interaction between abstract group theory and the theory of graph automorphisms, leading to the construction of graphs with remarkable properties as well as to a better understanding and occasionally a construction or proof of nonexistence of certain finite simple groups. On the other hand, in contrast to classical geometries, most finite graphs have no automorphisms other than the identity (asymmetric graphs), a fact that is largely and somewhat paradoxically responsible for its seeming opposite: every (finite) group is isomorphic to the automorphism group of a (finite) graph.

The study of graphs via their symmetries is rooted in the classical paradigm, stated in Felix Klein's "*Erlanger Programm*", that geometries are to be viewed as domains of a group action. Although graphs, as incidence structures, may seem to be degenerate geometries, we note that any incidence structure (such as a projective plane) can be represented by a graph. (The *Levi graph* L of an incidence structure S is a bipartite graph; its vertices correspond to the points and lines of S; and adjacency of vertices of L corresponds to point–line incidence in S.) Such representations preserve symmetry and allow fruitful generalizations (such as "generalized polygons", chapter 13, section 7).

In this chapter we try to illustrate the variety of ways in which groups and graphs interact. The effect of powerful results of group theory (such as the Feit–Thompson theorem on the solvability of groups of odd order) will be evident already in the introductory section 1.1. Consequences of the *classification of finite simple groups*

(CFSG) are required for some of the results in section 4.3 and for the analysis of some of the algorithms in sections 6.6 and 6.7. Many of the results surveyed in section 5 critically depend on the CFSG. On the other hand, some results of graph theoretic nature have played a role in the classification theory itself, as illustrated in sections 3.5 and 5.1.

In spite of these connections, *the treatment of the subject will mostly be kept on an elementary level*, requiring little more than basic group theory. The main theme of section 3 is the surprisingly strong effect of modest symmetry assumptions on the combinatorial parameters of a graph.

We try also to illustrate some of the links of the subject to areas not immediately seen to relate to groups. Sections 1.6 and 7.2 illustrate this point within combinatorics. Several connections to topology are explored in section 3 (see esp. sections 3.6 and 3.7). Random walks feature in section 3.8; linear algebra is visited briefly in sections 1.5, 3.8, 3.12 and 7.2.

Strong links have been forged to model theory (section 5.3) and to the theory of algorithms (sections 6.6 and 6.7). Some of the remote sources of motivation include algebraic topology (section 3.11), differential geometry (sections 3.6 and 3.7), and even quantum mechanics (section 1.5).

0.2. Isomorphisms, categories, reconstruction

Isomorphisms of graphs are bijections of the vertex sets preserving adjacency as well as non-adjacency. In the case of directed graphs, orientation must be preserved; in the case of graphs with colored edges and/or vertices, we agree that colors, too, must be preserved. Similar definitions apply to hypergraphs. In the case of incidence structures consisting of "points" and "lines", linked by incidence relations, we think of an isomorphism as a pair of bijections (one between the points, another between the lines), so that the pair preserves incidence. This view should be applied to graphs as well if multiple edges are allowed.

Automorphisms of the graph $X = (V, E)$ are $X \to X$ isomorphisms; they form the subgroup $\mathrm{Aut}(X)$ of the symmetric group $\mathrm{Sym}(V)$. Automorphisms of directed graphs, etc., are defined analogously.

The questions of *reconstruction* are, broadly speaking, questions of invertibility of certain isomorphism preserving operations on structures. A category in which all morphisms are isomorphisms is called a Brandt groupoid. Let $\mathscr{C}$, $\mathscr{D}$ be two Brandt groupoids and $F: \mathscr{C} \to \mathscr{D}$ a functor. Hence $X \cong Y$ implies $F(X) \cong F(Y)$. We call F *weakly reconstructible* if the converse also holds: $F(X) \cong F(Y)$ implies $X \cong Y$. We say that F is *strongly reconstructible* if for every pair X, Y of objects of $\mathscr{C}$, F induces a bijection between the sets $\mathrm{Iso}(X, Y)$ and $\mathrm{Iso}(F(X), F(Y))$ of isomorphisms. In this case, $\mathrm{Aut}(X) \cong \mathrm{Aut}(F(X))$ for every object X. We also say that, within the class $\mathscr{C}$, the object X is (weakly, strongly) reconstructible from $F(X)$.

A classical example is the reconstructibility of a multiset of direct irreducible finite groups from their direct product (unique direct factorization, R. Remak and O. Yu. Schmidt, cf. Baer 1947). The category $\mathscr{C}$ consists of the multisets of direct irreducible finite groups with the natural notion of isomorphism. Let F associate the direct product of the members of such a multiset X with X. This functor is weakly but not strongly reconstructible. (To see the latter, consider the pair $\{\mathbb{Z}_p, \mathbb{Z}_p\}$.)

Homomorphisms of graphs are defined as adjacency preserving maps, i.e., a map $f: V_1 \to V_2$ is a homomorphism of the graph $X_1 = (V_1, E_1)$ to the graph $X_2 = (V_2, E_2)$ if $(f(x), f(y)) \in E_2$ whenever $(x, y) \in E_1$. It is not required that non-adjacency be preserved; therefore a bijective homomorphism is not necessarily an isomorphism. It is easy to see that the chromatic number of the graph X is the smallest (cardinal) number m such that the set $\mathrm{Hom}(X, K_m)$ of $X \to K_m$ homomorphisms is nonempty. The set $\mathrm{End}(X) = \mathrm{Hom}(X, X)$ forms a monoid (semigroup with identity) under composition: the *endomorphism monoid* of X. $\mathrm{Aut}(X)$ consists of the invertible elements of $\mathrm{End}(X)$. The class of graphs together with the homomorphisms forms a *category*. These concepts extend naturally to directed graphs (orientation of edges must be preserved), graphs with colored vertices and/or edges (homomorphisms preserve color by definition); and to general relational structures involving relations of arbitrary arities.

The interconnections of these areas are manifold. The algorithmic problem of deciding whether or not two given graphs are isomorphic is equivalent to determining the automorphism group, and specific automorphism information for certain classes of graphs made it possible to use group theory to surprising depth in the analysis of graph isomorphism algorithms. Isomorphism rejection tools include graph invariants, i.e., functions F such that $X \cong Y$ implies $F(X) = F(Y)$. The construction of combinatorial, algebraic, and topological structures with prescribed automorphism groups and endomorphism monoids usually amounts to constructing strongly reconstructible functors. Reconstruction itself is an isomorphism problem, and automorphism groups have played a role in its study. Finally, establishing reconstructibility of certain functors is a useful tool in determining the automorphism groups of certain derived structures.

1. Definitions, examples

In this section, we collect some illustrative facts about automorphism groups of graphs and their interplay with reconstruction type problems.

We start with the simplest examples. A graph and its complement have the same automorphisms. The automorphism group of the complete graph K_n and the empty graph $\overline{K_n}$ is the symmetric group S_n, and these are the only graphs with doubly transitive automorphism groups. The automorphism group of the cycle of length n is the dihedral group D_n (of order $2n$); that of the directed cycle of length n is the cyclic group $\mathbb{Z}_n$ (of order n). A path of length $\geqslant 1$ has two automorphisms. The automorphism group of a graph is determined by the automorphism groups and the isomorphisms of its connected components: if $X_1, \ldots, X_k$ are pairwise nonisomorphic connected graphs, and X is the disjoint union of m_i copies of X_i, $i = 1, \ldots, k$, then

$$\operatorname{Aut}(X) = \operatorname{Aut}(X_1) \wr S_{m_1} \times \cdots \times \operatorname{Aut}(X_k) \wr S_{m_k}. \qquad (1)$$

The wreath products occurring here realize their imprimitive action (cf. chapter 12).

1.1. Measures of symmetry

A graph is *vertex-transitive* if its automorphism group acts transitively on the set of vertices. Such a graph is necessarily regular; the union of a 3-cycle and a 4-cycle show that the converse does not hold. If the group acts transitively on edges, the graph is *edge-transitive.* A vertex-transitive graph need not be edge-transitive. (Example: triangular prism.) If X is an edge-transitive graph without isolated vertices, and X is not vertex-transitive, then it must be bipartite, with the group acting transitively on each color class. The complete bipartite graphs $K_{m,n}$ with $m \neq n$ show that this can indeed happen. *Regular* graphs with edge- but not vertex-transitive automorphism groups are not so easy to construct (cf. Folkmann 1967, Bouwer 1969, 1972, Titov 1975, Klin 1981).

A *flag* in a graph X is an ordered pair (v, e) where v is a vertex and e is an edge incident with v. If $\operatorname{Aut}(X)$ is transitive on flags then X is *flag-transitive.* This means transitivity on the set of ordered pairs of adjacent vertices. For graphs without isolated vertices, flag-transitivity implies both vertex- and edge-transitivity. Again, the converse is false (cf. Holt 1981, Cameron 1983). If, however, X has *odd degree*, then vertex- and edge-transitivity imply flag-transitivity.

A graph X is *vertex-primitive* if $\operatorname{Aut} X$ is a primitive group. Vertex-primitivity by definition implies vertex-transitivity, but it does not imply edge-transitivity. (Take a cycle of prime length $p \geqslant 7$ and add all chords of length 2). For a graph X, let $X^{(t)}$ denote the graph obtained by joining a pair of vertices of X if their distance in X is t. If X is vertex-primitive and not empty then $X^{(t)}$ is connected for every $t \leqslant \operatorname{diam}(X)$. In particular, nonempty bipartite graphs X of order $\geqslant 3$ are never vertex-primitive (since $X^{(2)}$ is disconnected).

A graph is *distance-transitive* if $\operatorname{Aut}(X)$ is transitive on the set of ordered pairs of vertices at distance t for every $t \leqslant \operatorname{diam}(X)$. Nice examples are the Platonic solids (fig. 1 in chapter 1, section 1), Heawood's, Petersen's, and Coxeter's graphs (figs. 4, 8, 9 in chapter 1, sections 1 and 4).

Vertex-primitivity is a very severe restriction on the automorphism group, as seen by the following deep result previously known as the "Sims conjecture (1967)".

Theorem 1.1 (Cameron, Praeger, Saxl and Seitz 1983). *There exists a function f such that if a vertex-primitive digraph has out-degree k then the vertex-stabilizer in the automorphism group has order $\leqslant f(k)$.*

This result immediately implies that there is only a finite number of vertex-primitive distance-transitive graphs of any fixed degree. However, this second statement remains valid even without the vertex-primitivity condition (see section 5.2).

The automorphism group of a finite *tournament* T has *odd order*, since otherwise it would contain an involution (an element of order two), which would then illegally reverse at least one edge. This harmless looking observation implies, by the Feit–Thompson theorem, that $\mathrm{Aut}(T)$ is *solvable*, a fact with far reaching consequences, including algorithmic ones (cf. the end of section 6.6). Here we state an immediate corollary (cf. chapter 12 for the definitions).

Proposition 1.2. *Let T be a tournament with n vertices.*

(a) *If T is vertex-transitive then n is odd.*

(b) *If T is vertex-primitive then n is an odd prime power.*

Proof. Part (a) is straightforward: the in- and out-degrees must be equal. As for part (b), let N be a minimal normal subgroup of $\mathrm{Aut}(T)$. Then N is transitive (since $\mathrm{Aut}(T)$ is primitive); it is abelian (since $\mathrm{Aut}(T)$ is solvable); and it is characteristically simple (i.e. the direct product of isomorphic simple groups) (since it is minimal). Therefore $N \cong \mathbb{Z}_p^k$ ($k \geqslant 1$, p prime). A transitive abelian group being regular, we conclude that $n = |N| = p^k$. (We note that T is a Cayley digraph of N.) □

While there is no hope to classify all flag-transitive graphs, a simple description of all edge-transitive tournaments exists. (For directed graphs, edge- and flag-transitivity mean the same.) Let $q = p^k$ be an odd prime power, $q \equiv -1 \pmod 4$. The *Paley tournament* $P(q)$ has the field $\mathrm{GF}(q)$ for its vertex set; an edge goes from x to y ($x \neq y$) if $x - y$ is a square. The group of affine transformations $x \mapsto ax + b$ ($a, b \in \mathrm{GF}(q), a \neq 0$ a square) acts transitively on the edges of $P(q)$.

Theorem 1.3 (Kantor 1969).

(a) *Every edge-transitive tournament with $n > 2$ vertices is Paley.*

(b) $\mathrm{Aut}(P(q))$ *consists of the affine semilinear transformations $x \mapsto ax^{\alpha} + b$ where $a, b \in \mathrm{GF}(q)$, $a \neq 0$ is a square, and $\alpha : x \mapsto x^{p^j}$ $(0 \leqslant j \leqslant k - 1)$ is an automorphism of* $\mathrm{GF}(q)$.

Proof. Let T be edge-transitive. Since $n > 2$, T must be vertex primitive and therefore $n = p^k$, p an odd prime. The stabilizer of a vertex x acts transitively on the tournament induced by the $(n-1)/2$ out-neighbors of x; hence $n \equiv -1 \pmod 4$. Let N be a minimal normal subgroup of $G = \mathrm{Aut}(T)$; then, as before, N can be identified with the vertex set of T. Let $\tau : x \mapsto x^{-1}$ ($x \in N$); then τ is an *antiautomorphism* of T (reverses every edge). Therefore G has index 2 in the doubly transitive group $H = \langle G, \tau \rangle$. All solvable doubly transitive groups have been determined by Huppert (cf. Huppert 1957); apart from a finite number of exceptions of degrees $3^2, 5^2, 7^2, 11^2, 23^2, 3^4$, they all are subgroups of the group $\Gamma A_1(p^k)$ of semiaffine (affine semilinear) transformations of $\mathrm{GF}(p^k)$. The exceptional cases are ruled out because k must be odd (since $n \equiv -1 \pmod 4$). This, in particular, proves part (b). $\mathrm{Aut}(P(p^k))$ is the unique subgroup of index 2 in $\Gamma A_1(p^k)$, hence $G \leqslant \mathrm{Aut}(P(p^k))$. Since both G and $\mathrm{Aut}(P(p^k))$ have rank 3 (cf. chapter 12), either T or its converse agrees with $P(p^k)$, which is self-converse. □

The rth *residue digraph* $P(q,r)$ is defined for prime powers q and integers $r \geqslant 2$ such that $r|(q-1)$. The vertex set of $P(q,r)$ is GF(q); an edge joins x to y if $x-y$ is a rth power in GF(q). This digraph is undirected if either q or $(q-1)/r$ is even. (The *Paley graphs* are the quadratic residue graphs ($r=2$; $q \equiv 1 \pmod 4$)). The *Clebsch graph* is $P(16,3)$.) The affine linear group $A_1(q)$ is flag-transitive on $P(q,r)$. It is not true in general that Aut($P(q,r)$) is semiaffine; e.g., if $q=q_0^2$ and $r=q_0+1$ then $P(q,r)$ is a disjoint union of cliques; if $q=q_0^4$ and $r=q_0+1$, then the neighbors of 0 form a quadric and the graph admits the orthogonal group. However, the Paley graphs have semiaffine automorphism groups. This is a consequence of the following theorem of Carlitz (1960) and McConnel (1963): *Let q be a prime power, $r|q-1$, and let f be a map of* GF(q) *to itself such that for every $x,y \in$* GF(q), *$x \neq y$, the element $(x-y)^{-1}(f(x)-f(y))$ is an* rth *power. Then f is semiaffine.* (See also Bruen and Levinger 1973b.)

A stronger result holds when q is a prime.

Theorem 1.4. *If X is an edge-transitive regular graph of prime order p without isolated vertices then X is either complete or an* rth *residue graph for some $r|(p-1)/2$. In the latter case,* $\mathrm{Aut}(X)=A_1(p)$.

Proof. X cannot be bipartite (p is odd), hence it is vertex-transitive and (being a Cayley graph of the abelian group $\mathbb{Z}_p$, cf. Corollary 3.6), in fact, flag-transitive. Let $G=\mathrm{Aut}(X) \leqslant S_p$. If G is not solvable, then it is doubly transitive (cf. Burnside 1911; cf. also Huppert 1967, p. 609), hence X is complete. If G is solvable then $G \leqslant A_1(p)$ (Galois; see Huppert 1967, p.163). A glance at the structure of $A_1(p)$ completes the proof. □

Graphs with higher degrees of symmetry will be discussed in section 5. *Distance-transitive* graphs have been defined above. We define another important class here.

An s-arc in a graph is a sequence $(x_0,\ldots,x_s)$ of vertices such that: (a) x_{i-1} and x_i are adjacent; (b) $x_{i-1} \neq x_{i+1}$. The graph X is *s-arc-transitive*, if Aut(X) acts transitively on the set of s-arcs. (Note: 1-arc-transitivity is the same as flag-transitivity.) Distance-transitivity implies $\lfloor g/2 \rfloor$-arc-transitivity, where g is the *girth*.

Often we are interested in the action of some subgroup $G \leqslant \mathrm{Aut}(X)$ on vertices, edges, flags, etc. If this action is transitive (regular), we say that G is vertex-transitive (vertex-regular, resp.), etc., on X.

Graphs with relatively low degrees of symmetry are easy to construct. Every Cayley graph (see section 2) is vertex transitive. There is an abundance of edge-transitive digraphs and even of 2-arc-transitive graphs, as indicated by the following result. A map $f:(V,E) \to (W,F)$ between two finite digraphs is a k-fold *covering* if f is a homomorphism (maps vertices to vertices, edges to edges, and preserves incidences); every vertex and edge of (W,F) has exactly k preimages; and f is a local isomorphism, i.e. x and $f(x)$ have the same indegree (out-degree, resp.) for every $x \in V$.

Theorem 1.5 (Babai 1985). (a) *Every finite regular digraph has infinitely many edge-transitive finite covering digraphs with the same number of connected components.*

(b) *Every finite regular graph has infinitely many 2-arc-transitive finite covering graphs with the same number of connected components.*

It follows by a result of Godsil (1982) that *the minimal polynomial of every digraph divides that of an edge-transitive digraph,* hence the adjacency matrices of infinitely many edge-transitive digraphs are not diagonalizable.

Although graphs with higher symmetry are much more difficult to construct (cf. section 4), covering graphs are helpful in moving from an isolated example to infinitely many.

Theorem 1.6 (Biggs 1974, chapter 19). *A finite connected s-arc-transitive graph has infinitely many finite connected s-arc-transitive covering graphs.*

1.2. Reconstruction from line graphs

We illustrate the point made in the last sentence of the introduction by a classical example.

Theorem 1.7 (Whitney 1932). *Connected graphs X with $\geqslant 5$ vertices are strongly reconstructible from their line graphs $L(X)$ (within the class of all graphs).*

(Whitney proved the result for finite graphs; it was extended to infinite graphs by Bednarek (1985), using Rado's selection principle (chapter 42, section 3).) In other words, every isomorphism $L(X) \to L(Y)$ is induced in the natural way by a unique isomorphism $X \to Y$ (cf. Lovász 1979a, p.507). This, in particular, means that if the connected graph X has at least 5 vertices then $\mathrm{Aut}(X) \cong \mathrm{Aut}(L(X))$.

Corollary 1.8. *Let P denote the Petersen graph.*

(a) $\mathrm{Aut}(P) \cong S_5$.

(b) *P is distance transitive and 3-arc-transitive.*

Proof. The complement of P is $L(K_5)$. □

One can generalize this result to the *Kneser graphs* $\mathrm{KG}(n,r)$ $(n \geqslant 2r+1)$. Recall that the vertex set of $\mathrm{KG}(n,r)$ is the set of r-subsets of an n-set; disjoint subsets correspond to adjacent vertices.

Proposition 1.9.

(a) *For $n \geqslant 2r+1$,* $\mathrm{Aut}(\mathrm{KG}(n,r)) \cong S_n$.

(b) $\mathrm{KG}(n,r)$ *is distance-transitive.*

(c) *The "odd graph" $O_k = \mathrm{KG}(2k-1, k-1)$ is exactly 3-arc-transitive.*

For the proof of part (a), we have to consider a reconstruction problem for hypergraphs. The line graph $L(H)$ of the hypergraph $H = (V,E)$ has vertex set E; two members of E are adjacent in $L(H)$ if they intersect. For a set A, let $[A]^r$ denote the complete r-uniform hypergraph on A, consisting of all r-subsets of A. The Kneser graph $\mathrm{KG}(n,r)$ is the complement of $L([A]^r)$ where $|A| = n$. Part (a) of Proposition 1.9 is thus an immediate consequence of the next observation.

Proposition 1.10 (Berge 1972, Fournier 1974). *The complete r-uniform hypergraphs with at least $2r+1$ vertices are strongly reconstructible from their line graphs.*

Proof. By the Erdős–Ko–Rado theorem (see chapter 24), the largest cliques of $L([A]^r)$ are in one-to-one correspondence with the elements of A. This guarantees that every isomorphism $L([A]^r) \to L([B]^r)$ is induced by a bijection $A \to B$. □

This is a special case of the following sufficient condition of reconstructibility.

Theorem 1.11 (Erdős and Füredi 1980). *Let H be an r-uniform hypergraph on n vertices. If $n \geqslant 2r+1$ and every vertex of H has degree greater than*

$$v(n,r) = \binom{n-1}{r-1} - \binom{n-r-1}{r-1} + 1,$$

then H is strongly reconstructible from $L(H)$.

The degree bound $v(n,r)$ is tight for every $r \geqslant 2$ and $n > 2r^2$. The quantity $v(n,r)$ comes from the Hilton–Milner theorem (chapter 24, Theorem 5.8). In the particular case when all pairs of edges intersect in at most one point, the bound of Theorem 1.11 can be greatly improved.

Theorem 1.12. *Let H be an r-uniform hypergraph on n vertices such that every pair of edges intersects in at most one point. If every vertex of H has degree greater than $r^2 - r + 1$, then H is strongly reconstructible from $L(H)$.*

The proof follows immediately from Deza's theorem (1973) (cf. Lovász 1979a, problem 13.17): If every pair of edges of an r-uniform hypergraph $H = (V, E)$ has exactly λ points in common then either H is a sunflower (all edges have the same λ points in common), or $|E| \leqslant r^2 - r + 1$.

Corollary 1.13. *Let S, S_1 and S_2 be Steiner triple systems of order $\geqslant 15$. Then:*
(a) $\mathrm{Aut}(L(S)) \cong \mathrm{Aut}(S)$.
(b) *If $S_1 \ncong S_2$ then $L(S_1) \ncong L(S_2)$.*

We shall use part (a) of this corollary to construct strongly regular graphs with arbitrary prescribed automorphism groups (Theorem 4.3). Part (b) implies the existence of a large number of *isospectral graphs*: nonisomorphic graphs with the same characteristic polynomial. The existence of such families shows that the characteristic polynomial, though a useful invariant of graphs, is far from complete. (A *complete* invariant $F(X)$ is one from which X is (weakly) reconstructible.)

Corollary 1.14. *For infinitely many values of n, there exists a set of $n^{n(1+o(1))/2}$ isospectral graphs on n vertices.*

Indeed, the parameters of the strongly regular graph $L(S)$ (chapter 15) and therefore its spectrum are uniquely determined by the number of vertices $n = v(v-1)/6$, where v is the number of vertices of the Steiner triple system S. The estimate of the number of Steiner triple systems required, $v^{v^2(1+o(1))/6}$, is due to Alekseiev (1974) and Wilson (1974), combined with Van der Waerden's permanent conjecture (now the theorem of Egorychev and Falikman, see chapter 22, section 16.1).

A more direct proof of Corollary 1.14 (also based on the Permanent conjecture) uses Latin square graphs (LSGs). The LSG associated with a $k \times k$ Latin square (LS) (chapter 14) has k^2 vertices corresponding to the cells of the Latin square; two cells are adjacent in the graph if they are in the same row, or in the same column, or they have the same entry. For $k \geqslant 5$, the only k-cliques in an LSG are those corresponding to rows, columns, and identical entries. From this it is easy to deduce that (for $k \geqslant 5$) the LS is strongly reconstructible from its LSG. (Isomorphisms of Latin squares have to be defined carefully: row indices, column indices, and entries play interchangeable roles; so the automorphism group is a subgroup of $S_k \wr S_3$.)

1.3. Automorphism groups: reduction to 3-connected graphs

Probably the first nontrivial class of graphs of which the automorphism groups have been studied are finite trees (Jordan 1869). The first observation is that every tree has a *center*, which is either a vertex or an edge and is fixed under every automorphism. This reduces the problem to rooted trees (the root is fixed by definition). Automorphism groups of rooted trees can be determined recursively: delete the root, designate its neighbors to be roots of the remaining branches, and apply formula (1) to the forest of rooted trees obtained. The conclusion is as follows.

Proposition 1.15 (Jordan 1869). *The finite group G is isomorphic to the automorphism group of a finite tree if and only if $G \in \mathcal{W}$, where the class $\mathcal{W}$ of finite groups is defined inductively as follows:*

(a) $\{1\} \in \mathcal{W}$;

(b) *if $G, H \in \mathcal{W}$ then $G \times H \in \mathcal{W}$;*

(c) *if $G \in \mathcal{W}$ and $m \geqslant 2$ then $G \wr S_m \in \mathcal{W}$.*

In fact, not only the abstract group structure but the permutation action of the automorphism groups of trees can be deduced from these considerations.

Using the block-cutpoint tree T of a 1-connected graph X, similar considerations reduce the determination of $\mathrm{Aut}(X)$ to the automorphism groups of its blocks via a slight generalization of wreath products. If the root of T is a cutpoint, we split it and combine, via eq. (1), the groups of the (rooted) components. If the root is a block, we assign colors to the vertices of that block to indicate the isomorphism type of the incident branch; apply an arbitrary color preserving automorphism to the block, and move the branches in a wreath-product-like fashion (Robinson 1970).

A canonical decomposition of 2-connected graphs to their 3-connected "components" also exists.

We briefly indicate the idea. Let us call a multigraph *basic* if it is either 3-connected or a cycle or it has just two vertices and a set of $\geqslant 2$ parallel edges between them. A "bipolar multigraph" is a multigraph with two distinct specified endpoints. Call a bipolar multigraph basic if it becomes a basic multigraph after adding a new edge joining the two endpoints.

Let us now take a basic graph, and repeat the following construction: simultaneously replace every edge by a basic bipolar multigraph.

The result is that every 2-connected graph arises in a canonical way in this manner. Canonicity means that all isomorphisms between two 2-connected graphs induce isomorphisms of each corresponding level of this construction (and in particular it induces an isomorphism of the rooted trees representing the hierarchy of the basic graphs used).

Such a canonical hierarchy of basic graphs is referred to as the decomposition to 3-connected components.

A generalization of wreath products (Babai 1975) allows a description of the automorphism group of a 2-connected graph in terms of the automorphism groups of its 3-connected components with the edges of these components colored and oriented appropriately.

A very efficient (linear time) algorithm for the canonical decomposition to 3-connected components was given by Hopcroft and Tarjan using breadth-first search (Hopcroft and Tarjan 1973); a parallel algorithm was found by Miller and Ramachandran (1992).

Problems of great depth arise in the study of the automorphism groups of infinite trees. Tits (1970) studied the full automorphism groups of (vertex-colored) trees. Groups acting on trees without inverting an edge have been characterized by H. Bass and J.-P. Serre. This theory will be touched upon in sections 3.7 and 3.11.

1.4. Automorphism groups of planar graphs

Finite planar graphs form one of the few comparatively rich classes of graphs of which the automorphism groups have been satisfactorily determined, both from the algebraic (Babai 1975) and the algorithmic (Hopcroft and Tarjan 1972; Hopcroft and Wong 1974) points of view.

Every finite group of isometries of the euclidean 3-space has a fixed point and can therefore be identified with a group of isometries of the 2-sphere. Every sense-preserving transformation is a *rotation,* and every sense-reversing transformation is a *rotary inversion*, i.e. a rotation followed by a central inversion.

There are two infinite families and 3 sporadic examples of finite *rotation groups* of the 2-sphere: the rotation groups of the regular k-gonal pyramids (the cyclic group $\mathbb{Z}_k$), the regular k-gonal prisms (the dihedral group D_k), the tetrahedron (the alternating group A_4), the cube (S_4), and the dodecahedron (A_5) (see the figures in chapter 1, section 1). The list is understood to include the degenerate cases $k \leqslant 2$.

The finite *isometry groups* of the 2-sphere, other than the rotation groups, can be obtained in one of two ways as follows. Each rotation group G can be extended to $G \cup G\tau \cong G \times \mathbb{Z}_2$ where $\tau = -I$ is the central inversion. Moreover, if G is a rotation group with a subgroup H of index 2, then the group $G^* = H \cup (G \setminus H)\tau$ is another isometry group. Note that $G^* \cong G$, but the geometric realization is different: for instance, from the rotation group of the cube we obtain the full isometry group of the tetrahedron. (See, e.g., Fejes-Tóth 1965, Coxeter 1961.)

Theorem 1.16. *Every 3-connected planar graph X has an embedding on the sphere such that all automorphisms are realized by isometries of the sphere.*

This is a consequence of Whitney's (1932) theorem that 3-connected planar graphs are uniquely embeddable on the sphere (cf. chapter 2), combined with the fact that all finite homeomorphism groups of the sphere are topologically equivalent to a group of isometries (Kerékjártó 1921, Eilenberg 1934). A stronger version of Theorem 1.16 was obtained by P. Mani.

Theorem 1.17 (Mani 1971). *Every 3-connected planar graph X can be realized as the 1-skeleton of a convex polytope P in $\mathbb{R}^3$ such that all automorphisms of X are induced by isometries of P.*

Polyhedral groups are the isometry groups of convex polytopes and their subgroups. Viewed in their action on $\mathbb{R}^3$, they coincide with the finite isometry groups listed above. Either of the above results, combined with the reduction process indicated in the previous section, yields a description of the automorphism groups of planar graphs in terms of generalized wreath products of symmetric groups and polyhedral groups. Two easily stated consequences: If X is planar then Aut X has a subnormal chain $\mathrm{Aut}(X) = G_0 \triangleright G_1 \triangleright \cdots \triangleright G_m = 1$ such that each quotient group G_{i-1}/G_i is either cyclic or symmetric or A_5. If X is 2-connected and $|\mathrm{Aut}(X)|$ is odd then $\mathrm{Aut}(X)$ is cyclic (Babai 1975). We conjecture that the first of these statements remains valid for graphs embeddable on an arbitrary fixed surface Σ (cf. chapter 5) with A_5 replaced by a finite list, depending on Σ (cf. Babai 1973, 1974a).

1.5. Matrix representation. Eigenvalue multiplicity

A mechanical system is often represented by a self-adjoint operator A; and its symmetries by a group G of unitary operators (acting on a real or complex Hilbert space H). The fact of symmetry is expressed by the equation $AP = PA$ for each $P \in G$. If H has finite dimension (or more generally, its spectrum is discrete), then it is the orthogonal direct sum of the eigensubspaces $H_\lambda = \{u \in H\colon Au = \lambda u\}$ for all eigenvalues λ of A.

If the operators B and C commute, then the eigensubspaces of B are invariant subspaces for C. In particular, one can refine the decomposition $H = \sum^{\oplus} H_\lambda$ to an orthogonal decomposition into subspaces, irreducible under the action of G. This way each irreducible constituent of G falls into an eigensubspace, forcing "degeneracies" (multiple eigenvalues) to occur, and more importantly, the vectors in an orthonormal base of H are classified according to the irreducible constituents of G. This approach, introduced in a seminal paper by Wigner (1927) has been used extensively both in classical and in quantum mechanics (cf. Wigner 1959, Hamermesh 1962). The classification of eigenvibrations of molecules using the character tables of their symmetry groups (also due to Wigner 1930, cf. Schonland 1965) is particularly instructive because in this case $\dim H < \infty$ and the matrix A is a variant of the "adjacency matrix of the molecule".

Let now X denote a graph with edges weighted with real numbers; and let A be its adjacency matrix; so the entry $a_{i,j}$ is the weight of the edge $\{i,j\}$. Then A is a symmetric real matrix which acts on the space $H = \mathbb{R}^n$ (n is the number of vertices). The automorphisms of A are represented by precisely those permutation matrices P which commute with A.

Reversing Wigner's approach, we shall indicate how to use spectral information on A to infer properties of the group $G = \mathrm{Aut}(X)$. Let $\lambda_1, \ldots, \lambda_m$ be the eigenvalues of A; let m_i be the multiplicity of λ_i ($\sum m_i = n$). Let G_i denote the restriction of G to the eigensubspace H_{λ_i}. Then G is a subdirect product of the G_i. (A *subdirect product* is a subgroup of the direct product which projects *onto* each factor.) This proves part (a) of the following result.

Theorem 1.18. *Let* $G = \mathrm{Aut}(X)$ *for an edge-weighted graph* X *with eigenvalue multiplicities* $m_1 \leqslant \cdots \leqslant m_t$.

(a) (Godsil 1978) *G is the subdirect product of groups $G_1, \ldots, G_t$, where G_i is a subgroup of the orthogonal group* $\mathrm{O}(m_i)$;

(b) (Godsil 1978) $|G_i| \leqslant n^{m_i}$;

(c) (Godsil 1978) *if X is vertex-primitive then* $|G| \leqslant n^{m_2}$;

(d) (Babai, Grigoryev and Mount 1982) *if X is vertex-transitive then* $|G| \leqslant n^{m_t - 1}$; *and more generally, the restriction of G to any of its orbits has order* $\leqslant n^{m_t - 1}$.

To see part (b), let S be the projection of the standard basis of $H = \mathbb{R}^n$ to H_{λ_i}; and let $S' \subseteq S$ be a base of H_{λ_i}. Then each member of G_i is determined by its restriction to S' which is a map $S' \to S$. The number of such maps is $\leqslant n^{m_i}$. Part (c) follows by observing that the projection of V to each eigensubspace defines an invariant partition of V. Hence if X is vertex-primitive of degree $d \geqslant 1$ then this partition must be trivial and G acts faithfully on each eigensubspace of dimension $\neq 1$. But the only one-dimensional eigensubspace of a vertex-transitive graph is the one corresponding to $\lambda = d$. Part (d) is less immediate; an algorithmic version of it is used in Babai et al. (1982) to deduce an $n^{m+O(1)}$ algorithm for testing isomorphism of graphs with eigenvalue multiplicity bounded by m.

Part (a), too, has some appealing consequences. Let $m = m_t$ be the maximum multiplicity of eigenvalues. Noting that $\mathrm{O}(1) \cong \mathbb{Z}_2$, we see that *if all eigenvalues of a graph are distinct, then its automorphism group is an elementary abelian 2-group* (Mowshowitz, Petersdorf and Sachs 1969, cf. Cvetković et al. 1980). Further, *if* $m \leqslant 3$ *then* $\mathrm{Aut}(X)$ *is solvable.* This is immediate for $m = 2$ since every finite subgroup of $\mathrm{O}(2)$ is cyclic or dihedral; but among the finite subgroups of $\mathrm{O}(3)$, there are two nonsolvable ones: the group of rotations of the icosahedron ($\cong A_5$) and its full group of congruences ($\cong A_5 \times \mathbb{Z}_2$). These were ruled out by Cameron (1983) via a closer look at the characters of A_5.

Our last remark concerns factors of the characteristic polynomial. Let $G \leqslant \mathrm{Aut}(X)$ for some weighted digraph X and consider the weighted quotient graph $Y = X/G$. The vertices of X/G are the orbits of G; the weight of the directed edge (A,B) of Y is the sum of weights of all edges (u,v) for some fixed $u \in A$ over all $v \in B$. It is easy to see that the characteristic polynomial of Y divides that of X. In particular, *if the characteristic polynomial of a digraph X is irreducible then X is asymmetric* ($|\mathrm{Aut}(X)| = 1$) (Mowshowitz, cf. Cvetković et al. 1980).

1.6. Asymmetry, rigidity. Almost all graphs. Unlabelled counting

An excellent exposition of the subject of this section is given by Bollobás (1985, chapter IX).

A graph is called *asymmetric* if it has no nontrivial automorphisms; it is called *rigid* if it has no nontrivial endomorphisms. (Some authors use the term "rigid" to describe what we call asymmetric.) Construction of asymmetric or rigid graphs and other structures with given properties is often the basis of the construction of such structures with given automorphism group or endomorphism monoid, resp. (Cf. section 4.1.) A notable result in this area is that *there exists a rigid graph on every infinite vertex set* (Vopenka et al. 1965, cf. Hedrlín and Lambek 1969). Finite rigid graphs exist on n vertices for any $n \geqslant 10$ (Hedrlín and Pultr 1965); asymmetric graphs exist for $n \geqslant 6$. Asymmetry/rigidity is actually the typical behavior of finite graphs. It was proved by Pólya (1937), Erdős and Rényi (1963) that *a random graph is asymmetric* with probability $1 - \binom{n}{2}2^{-n-2}(1 + \mathrm{o}(1))$. The dominant part of the error-term comes from the graphs which admit a transposition automorphism (a pair P of vertices with identical neighborhood outside P). The asymptotic expansion can be continued to include terms describing the probabilities of automorphisms with bounded supports. A strong algorithmic version of this result will be mentioned in section 6.4.

It is not difficult to upgrade the proof to yield that almost all graphs are rigid. In this case the error term is $\mathrm{O}(n^2(3/4)^{-n})$, dominated by the possibility that the neigborhood of some vertex v includes the neighborhood of some vertex w, allowing an endomorphism to map w to v while fixing all other vertices.

Although n-vertex asymmetric trees exists for every $n \geqslant 7$, random trees are typically not asymmetric. Indeed for any finite rooted tree T, almost all labeled trees have T as a limb (Schwenk 1973). In particular, large numbers of cherries (pairs of end-vertices with a common neighbor) occur almost always.

Nontrivial trees (and more generally, bipartite graphs, and indeed perfect graphs) are never rigid (they can be mapped to their largest clique).

E. M. Wright refined the "almost sure asymmetry" results to show that asymmetry is typical for graphs with density above the *connectedness threshold* (cf. chapter 6).

Theorem 1.19 (Wright 1971). *Let $m(n) = \frac{1}{2}n \ln n + n\psi(n)$. Then the probability that a random graph with n vertices and $m(n)$ edges is asymmetric tends to* 1 *if $\psi(n) \to \infty$ assuming $m(n) \leqslant \frac{1}{2}\binom{n}{2}$; and this probability tends to* 0 *if $\psi(n) \to -\infty$.*

The reason of the second statement is obvious: those graphs have, with probability approaching 1, an unbounded number of isolated vertices. If we rule out this possibility, even sparser graphs will be typically asymmetric: for fixed $r \geqslant 3$, the probability that a random r-regular graph is asymmetric tends to 1 (Bollobás 1982, McKay and Wormald 1984, Wormald 1986).

The results establishing "almost always asymmetry" mentioned above are valid for labeled as well as for unlabeled graphs; the latter is a substantially stronger statement with important consequences to counting unlabeled objects. We shall formalize the connection below.

Let $\mathscr{C}$ be a class of finite graphs (or digraphs, or other structures), closed under isomorphisms, and let $\mathscr{C}(n)$ be the set of those members of $\mathscr{C}$ with vertex set $[n] = \{1, \ldots, n\}$. Let $\mathscr{P}$ be a graph property (i.e. an isomorphism-closed class of graphs). We say that "almost all labeled members of $\mathscr{C}$ have property $\mathscr{P}$" if $\lim_{n\to\infty} |\mathscr{P} \cap \mathscr{C}(n)|/|\mathscr{C}(n)| = 1$. The term "almost all unlabeled members of $\mathscr{C}$" is used analogously except that isomorphism classes rather than individual graphs are counted. This annoying distinction disappears if almost all *unlabeled* members of $\mathscr{C}$ are asymmetric: under this condition, any graph property will hold for almost all unlabeled members of $\mathscr{C}$ if and only if it holds for almost all *labeled* members.

The statement that "almost all unlabeled members of $\mathscr{C}$ are asymmetric" is equivalent to the following:

"the expected number of automorphisms of a random labeled member of $\mathscr{C}$ is $1 + o(1)$." (2)

This equivalence follows from the observation that the *number of unlabeled graphs* (isomorphism classes) in $\mathscr{C}(n)$ is exactly $|\mathscr{C}(n)|\alpha(n)/n!$, where $\alpha(n) = \sum_{X\in\mathscr{C}(n)} |\mathrm{Aut}(X)|/|\mathscr{C}(n)|$ is the expected order of the automorphism group of a random labeled member of $\mathscr{C}$. (This follows from the Orbit counting lemma, a.k.a. "Burnside's lemma", see chapter 21, Lemma 14.3.)

By the results mentioned, (2) is valid for the class of all graphs, for graphs with $m(n)$ edges as in Wright's theorem ($\psi \to \infty$), as well as for regular graphs of given degree $r \geqslant 3$.

Structures satisfying stronger regularity constraints are often difficult to count. It seems likely, for instance, that almost all strongly regular graphs are asymmetric, but this may be difficult to prove. It has been shown, however, that almost all (unlabeled) members of the following two classes of strongly regular graphs are asymmetric: the line graphs of Steiner triple systems, and the Latin square graphs (cf. section 4.1) (Cameron 1979, Babai 1979a).

While almost all graphs are asymmetric, one might be interested in what can be said about the graphs known to admit some automorphisms. Related questions will be considered in sections 4.3 and 4.4; here we mention a result of Cameron (1980b).

Theorem 1.20 (Cameron). *For a finite group G let $\mathscr{C}(G)$ be the class of those graphs X admitting a group isomorphic to G as a subgroup of* $\mathrm{Aut}(X)$. *Let $a_n(G)$ denote the proportion of those n-vertex labeled members of $\mathscr{C}$ which have* $\mathrm{Aut}(X) \cong G$. *Then:*

(i) *the limit* $a(G) := \lim_{n\to\infty} a_n(G)$ *always exists and is rational.*

(ii) $a(G) = 1$ *iff* G *is the direct product of symmetric groups.*

(iii) *For infinitely many groups, including all abelian groups with exponent* $\geqslant 3$, $a(G) = 0$.

(iv) *For metabelian groups, the values of* $a(G)$ *are dense in* $[0,1]$.

While almost all finite graphs are asymmetric, the situation changes to its opposite when we consider countably infinite graphs. Let us generate a random graph on a countably infinite vertex set by deciding independently and with probability $\frac{1}{2}$ whether or not to join two vertices. Then with probability 1, we obtain a graph isomorphic to one specific graph R, the Rado graph (Erdős and Rényi 1963), discussed in section 5.3. We should mention that $|\mathrm{Aut}(R)| = 2^{\aleph_0}$, and "almost all" automorphisms of R are conjugates (cf. Theorem 3.17).

More generally, *the number of automorphisms of a countable graph* (or any countable structure over a locally finite language, cf. section 5.3) *is always finite, countable, or* $2^{\aleph_0}$. A countable graph (structure) X has $2^{\aleph_0}$ automorphisms if and only if every finite subset of $V(X)$ is pointwise fixed by some nontrivial automorphism.

2. Graph products

In this section we introduce the most important graph products, indicate their combinatorial significance, and address their automorphism and factoring problems.

Given two graphs $X_i = (V_i, E_i)$ $(i = 1,2)$, a *product graph* $Y = (W, F) = X_1 * X_2$ can be defined in a variety of sensible ways. Those four which appear most frequently in the literature are the lexicographic, the Cartesian, the categorical, and the strong products. In each case, $W = V_1 \times V_2$ (Cartesian product). Each of the products is associative, and three of the four are commutative in the sense that the map $(v_1, v_2) \mapsto (v_2, v_1)$ is an isomorphism between $X_1 * X_2$ and $X_2 * X_1$. (The lexicographic product is not commutative.) The 1-vertex graph is a (two-sided) identity in three cases (exception: the categorical product; in that case it is natural to admit loops and the one-vertex graph with a loop becomes the identity). We say that a graph P is a *prime* with respect to a product and a class $\mathscr{C}$ of graphs if P is not isomorphic to the product of two non-identity graphs within $\mathscr{C}$ and is not itself the identity.

Next we define the adjacency relation in each product. Let $u_i, v_i \in V_i$ and $(i = 1,2)$ $w = (u_1, u_2)$, $z = (v_1, v_2) \in W$. Then w and z are adjacent (a) in the *lexicographic product* $Y = X_1[X_2]$ if either $(u_1, v_1) \in E_1$ and $u_2 \neq v_2$, or $u_1 = v_1$ and $(u_2, v_2) \in E_2$; (b) in the *Cartesian product* $Y = X_1 \times X_2$ if either $u_1 = v_1$ and $(u_2, v_2) \in E_2$ or $u_2 = v_2$ and $(u_1, v_1) \in E_1$; (c) in the *categorical product* $Y = X_1 \cdot X_2$ if $(u_1, v_1) \in E_1$ and $(u_2, v_2) \in E_2$; (d) the edge set of the *strong product* $X_1 \otimes X_2$ is the union of the edge sets of the Cartesian and the categorical products.

Observe that the n-cube is the Cartesian product of n copies of K_2; more generally, Cartesian products of paths are grids. *Hamming graphs* can be defined as isometric subgraphs of Cartesian products of complete graphs (cf. Graham and Winkler 1985).

Categorical products are the products in the category theoretic sense. They give rise to some of the deepest structural questions (cf. McKenzie 1971, Jónsson 1981). Strong products, their close relatives, are tamer in many ways.

Lexicographic products occur naturally in combinatorial constructions; we shall mention examples below.

Some graph invariants of certain products are easily computed from those of the factors; others pose important open questions. We mention two of the latter kind. The first one concerns the *chromatic number* $\mathrm{chr}(Y)$ *of the categorical product* $Y = X_1 \cdot X_2$. Since Y has a homomorphism to each factor, clearly $\mathrm{chr}(Y) \leqslant \min\{\mathrm{chr}(X_1), \mathrm{chr}(X_2)\}$. Hedetniemi's conjecture asserts that for finite graphs we have equality here (cf. Greenwell and Lovász 1974). (This is false for uncountably infinite graphs, Hajnal 1985.)

The second problem concerns the *independence number* $\alpha(Y)$ *of the strong product* $Y = X_1 \otimes X_2$. Clearly $\alpha(Y) \geqslant \alpha(X_1)\alpha(X_2)$ (supermultiplicativity). Let $X^k = X \otimes \cdots \otimes X$ denote the kth strong power of the graph X. Supermultiplicativity implies that the limit $\Theta(X) = \lim_{k\to\infty}(\alpha(X^k))^{1/k}$ always exists; this quantity is the *Shannon capacity* of X (chapter 31, section 6; cf. Knuth 1994). Its value is unknown even for as simple a graph as C_7, the cycle of length 7. Even the case of C_5 was open for decades; it was solved by Lovász as a special case of the following result: *If X is a vertex-transitive self-complementary graph with n vertices then* $\Theta(X) = \sqrt{n}$ (Lovász 1979b). (This class includes the Paley graphs (section 1.1).)

Cartesian products of cycles occur as Cayley graphs of abelian groups. Their genus has been studied in this context (cf. section 3.9).

Some useful observations regarding the *lexicographic product:* (1) both the *independence number* $\alpha(X)$ and the *clique number* $\omega(X)$ are multiplicative under lexicographic products (this fact has a curious application to constructive Ramsey graphs, Abbott 1972). The following inequality holds for the chromatic number of the lexicographic product (Linial and Vazirani 1989):

$$(\mathrm{chr}(X_1) - 1) \cdot \mathrm{chr}(X_2)/\ln|V(X_1)| \leqslant \mathrm{chr}(X_1[X_2]) \leqslant \mathrm{chr}(X_1) \cdot \mathrm{chr}(X_2). \tag{3}$$

Sometimes the study of vertex-transitive graphs reduces to the study of Cayley graphs via the following observation: *If X is vertex-transitive then both $X[K_m]$ and $X[\overline{K}_m]$ are Cayley graphs for a suitable m* (Sabidussi 1964). (Examples include the study of isoperimetry, cf. Theorems 3.38, 3.41.)

Among the nicely behaved parameters we mention the *spectrum*. Let $\{\lambda_i\}$ and $\{\mu_j\}$ be the multisets of eigenvalues of X_1 and X_2, resp. Then $\{\mu_j + |X_2|\lambda_i\}$, $\{\lambda_i + \mu_j\}$, $\{\lambda_i\mu_j\}$, and $\{\lambda_i\mu_j + \lambda_i + \mu_j\}$ are the respective multisets of eigenvalues of the lexicographic, Cartesian, categorical, and strong products. All these products share a base of orthonormal eigenvectors consisting of the pairwise Kronecker products of the orthonormal eigenbases of each factor. The Kronecker product of the adjacency matrices of X_1 and X_2 is the adjacency matrix of their categorical product.

2.1. Prime factorization, automorphism group

Now we turn to the problem of *unique prime factorization* (UPF). We recommend the insightful survey by Imrich (1993) for more details.

For commutative products of finite graphs, UPF is equivalent to the *common refinement property*. We say that a graph G has the common refinement property with respect to a product, if for any two representations $\prod_{i\in I} A_i \cong \prod_{j\in J} B_j$ of G there exist graphs $C_{i,j}$ which satisfy $A_p \cong \prod_{j\in J} C_{p,j}$ and $B_q \cong \prod_{i\in I} C_{i,q}$.

Let $V = V_1 \times \cdots \times V_k$ be a Cartesian product decomposition of the vertex set V. For $v \in V$, let V_i^v denote the set of vertices differing from v in the ith coordinate only. Suppose this decomposition of V corresponds to a decomposition of the graph $X = (V, E)$ with respect to some commutative product; and $V = W_1 \times \cdots \times W_\ell$ corresponds to another decomposition. We say that the *strict common refinement property* (SCR) holds if the intersections $V_i^v \cap W_j^v$ with at least two vertices are exactly the $C_{i,j}^v$ with respect to the factors of a common refinement. We say that X has the SCR property w.r. to a certain product if any pair of product decompositions of X has this property. In this case it follows that the multiset of prime factors is *strongly reconstructible*. In particular, X has UPF $X = \prod X_i$ and $\mathrm{Aut}(X)$ is obtained from $\mathrm{Aut}(X_i)$ via eq. (1) at the beginning of section 1.

For disconnected graphs, UPF does not hold for any of our commutative products, as seen from the identity $(1+x+x^2)(1+x^3) = (1+x^2+x^4)(1+x)$. (Plug in a connected prime graph for x and interpret $+$ as disjoint union.)

2.2. The Cartesian product

The product of connected graphs is connected.

Every connected graph has UPF in a strong sense (Sabidussi 1960) which we now state. Every Cartesian product decomposition $X = Y_1 \times \cdots \times Y_k$ of the graph X induces an equivalence relation $\sigma(Y_1, \ldots, Y_k)$ on $E(G)$; equivalent edges correspond to edges of the same Y_i. It turns out that if X is connected then the intersection of two such *product relations* is a product relation again. The UPF corresponds to the intersection σ_X of all product relations. The strict common refinement property for connected graphs follows immediately, implying UPF and eq. (1) for the prime factors.

Several algorithms are known to construct the UPF. The simplest one is due to Feder (1992) and runs in $\mathrm{O}(mn)$ ($m = |E|$, $n = |V|$). The most efficient algorithm, found by Aurenhammer et al. (1992), runs in $\mathrm{O}(m \log n)$.

Unique prime factorization holds for infinite graphs as well, and extends to the *weak* Cartesian product of infinitely many connected graphs (Imrich). For this result and for the connections between prime factorization and *isometric embeddings* into Cartesian products we refer to Imrich (1989).

2.3. The categorical product; cancellation laws

First of all we have to admit loops so we at least have an identity graph for this product (single vertex with a loop). The categorical product of two connected graphs is bipartite iff at least one of them is bipartite; and it is disconnected iff both factors are bipartite. Disconnected products cause non-unique prime factorizations; but the connected non-bipartite graphs have UPF in the class of graphs with loops (McKenzie 1971). However, the strict refinement property does not hold, not even its consequence, eq. (1).

A graph is *thin* if no pair of vertices has precisely the same set of neighbors. All factors of connected, non-bipartite thin graphs have the same properties. *The strict common refinement property holds for connected, non-bipartite thin graphs*, with its usual consequences: UPF and eq. (1) for prime decomposition (McKenzie 1971).

The inference $A \cdot C \cong B \cdot C \Rightarrow A \cong B$ is called *cancellation*. The cancellation law is an immediate consequence of UPF; however, it may hold even if UPF fails. Lovász (1971) proved that for cancellation, it suffices to require that the finite graphs A and B both have a homomorphism to C. Moreover $A^n \cong B^n$ always implies $A \cong B$. In fact, Lovász (1972a) has shown, using an elegant inclusion–exclusion argument, that these statements hold *in any finite category*.

2.4. Strong product

For a simple graph X, let X_0 be the graph obtained by attaching a loop at each vertex. Let Y^0 be obtained by removing all loops from Y. Now for two simple graphs X, Y, we have $X \otimes Y = (X_0 \cdot Y_0)^0$. Thus the strong product can be viewed as a tame special case of the categorical product (imagine a loop at every vertex). It follows that for connected simple graphs, UPF holds. Moreover, the *strict common refinement property* holds for connected graphs with *thin complements*.

The UPF of connected graphs can be found in polynomial time (Feigenbaum and Schäffer 1992).

2.5. Lexicographic product

This product is *right-distributive* with respect to disjoint unions (all other products discussed are distributive). It distributes complementation: $\overline{X[Y]} \cong \overline{X}[\overline{Y}]$. The only pairs of graphs which *commute* with respect to the lexicographic product are (K_n, K_m), $(\overline{K}_n, \overline{K}_m)$, and X^n, X^m for any X. Moreover, the following *cancellation law* holds for finite graphs: if $A[B] \cong X[Y]$ and $|V(B)| = |V(Y)|$ then $A \cong X$ and $B \cong Y$.

Let $X + Y$ denote the disjoint union of the graphs X, Y and set $X \oplus Y = \overline{\overline{X} + \overline{Y}}$ (Zykov sum).

Observe that $\overline{K}_q[X[\overline{K}_q] + \overline{K}_m] \cong (\overline{K}_q[X] + \overline{K}_m)[\overline{K}_q]$, and, by complementation, $K_q[X[K_q] \oplus K_m] \cong (K_q[X] + K_m)[K_q]$. We call these operations *elementary transpositions*. They preserve primality.

Theorem 2.1 (Chang 1961, Imrich 1971). *Any two prime factorizations with respect to the lexicographic product can be transformed into each other by elementary transpositions.*

For further references, cf. Jónsson (1981).

Clearly, $\mathrm{Aut}(X[Y]) \leqslant \mathrm{Aut}(Y) \wr \mathrm{Aut}(X)$ (wreath product in its imprimitive action, cf. chapter 12): we may apply an automorphism of each copy of Y separately; and then, apply a single automorphism of X. We state a sufficient condition which guarantees equality here.

Theorem 2.2 (Sabidussi). *Let X, Y be finite graphs. Assume X is thin if Y is disconnected and $\overline{X}$ is thin if $\overline{Y}$ is disconnected. Then* $\mathrm{Aut}(X[Y]) = \mathrm{Aut}(Y) \wr \mathrm{Aut}(X)$.

Feigenbaum and Schäffer (1992) observed that recognizing composite graphs is polynomial-time equivalent to the graph isomorphism problem (section 6) and therefore not known to be solvable in polynomial time.

3. Cayley graphs and vertex-transitive graphs

3.1. Definition, symmetry

In 1878, Cayley introduced a graphic representation of abstract groups. With a group G and a set $S \subseteq G$ of generators he associated what we now call a *Cayley color diagram:* a directed graph with colored edges. The vertex set of the diagram $\Gamma_c(G,S)$ is G. A color corresponds to each member of S; and the vertex $g \in G$ is joined to $sg \in G$ by an edge of color s.

If we ignore colors, we obtain the *Cayley digraph* $\Gamma(G,S)$. If in addition we ignore orientation of the edges, we obtain a simple graph: the *Cayley graph* $\Gamma(G,S)$. The degree of its vertices is $|(S \cup S^{-1}) \setminus \{1\}|$.

The Cayley graph $\Gamma(G,S)$ is connected because S generates G. Cycles in the Cayley graph correspond to relations among the elements of S. In particular, if S is a set of free generators of a free group G then $\Gamma(G,S)$ is a tree. The converse also holds if there are no involutions (elements of order 2) in S. (Involutions correspond to cycles of length 2 in the Cayley diagram, invisible in the Cayley graph.) More generally, if $\Gamma(G,S)$ is a tree then G is a free product of infinite cyclic groups and of cyclic groups of order 2; the members of S generate these free factors.

If no proper subset of S generates G, we call $\Gamma(G,S)$ a *minimal Cayley graph*. Infinite groups do not normally have minimal sets of generators. If S can be linearly ordered such that no element of S is generated by its predecessors, we call $\Gamma(G,S)$ *semiminimal*. Every group possesses semiminimal Cayley graphs.

For $g \in G$, the *right translation* $\rho_g : G \to G$ is defined by $x\rho_g = xg$ $(x \in G)$. The map $\rho : g \mapsto \rho_g \in \mathrm{Sym}(G)$ is the *right regular permutation representation* of G. Its image $G\rho \leqslant \mathrm{Sym}(G)$ is a regular permutation group (chapter 12). The following statements regarding the automorphism groups of Cayley diagrams and graphs are easy to verify. (Recall that automorphisms of colored directed graphs preserve colors and orientation by definition.)

Proposition 3.1. (a) $G\rho = \mathrm{Aut}(\Gamma_c(G,S)) \leqslant \mathrm{Aut}(\Gamma(G,S))$.

(b) (Sabidussi 1964) *A connected graph $X = (G,E)$ is a Cayley graph of the group G if and only if $G\rho \leqslant$* $\mathrm{Aut}(X)$.

Cayley graphs are thus vertex-transitive; the converse of this statement is false. Indeed, by Proposition 3.1(b), a connected graph X is Cayley precisely if $\mathrm{Aut}(X)$ contains a regular subgroup. The smallest example of a vertex-transitive graph with no regular subgroup of automorphisms is Petersen's graph. This is the first member $\mathrm{KG}(5,2)$ of the infinite family of Kneser's graphs $\mathrm{KG}(n,r)(n \geqslant 2r+1 \geqslant 5)$, most of which are not even remotely Cayley-like. $\mathrm{KG}(n,r)$ has $\binom{n}{r}$ vertices identified with the set of r-tuples of an n-set; two vertices are adjacent if the corresponding r-tuples are disjoint (cf. chapters 4, 24, 34).

Theorem 3.2 (Kantor 1972, Godsil 1980a). (a) *Kneser's graph* $\mathrm{KG}(n,r)$ *is a Cayley graph precisely if* $r=2$ *and* n *is a prime power,* $n \equiv -1 \pmod 4$, *or* $r=3$ *and* $n \in \{8,32\}$.

(b) *If* $r \geqslant 4$ *then, with some exceptions, the only transitive proper subgroup of* $\mathrm{Aut}(\mathrm{KG}(n,r))$ *is the one induced by the alternating group* A_n. *Exceptions occur for* $r=5$ *when* $n \in \{12,24\}$ *and for* $r=4$ *when* $n \in \{9,11,12,23,24,33\}$.

The proof requires the following result of Livingstone and Wagner. A permutation group $G \leqslant \mathrm{Sym}(A)$ is t-transitive if it is transitive on the set of ordered t-tuples of distinct elements of A. G is t-homogeneous if it is transitive on the set of t-subsets of A.

Theorem 3.3 (Livingstone and Wagner 1965). (a) *If G is t-homogeneous then it is* $(t-1)$*-transitive.*

(b) *If G is t-homogeneous and* $t \geqslant 5$ *then G is t-transitive.*

Proof of Theorem 3.2. Assume that $\mathrm{KG}(n,r)$ is a Cayley graph of some group $G \leqslant \mathrm{Aut}(\mathrm{KG}(n,r))$. Then, by Proposition 1.9(a), we may view G as a subgroup of S_n. Now, G acts regularly on the r-subsets, and is therefore r-homogeneous. By Theorem 3.3, it must be r-transitive if $r \geqslant 5$. The case $r \geqslant 5$ now follows because of the nonexistence of nontrivial 4- and 5-transitive permutation groups of degrees other than those listed.

For $r \geqslant 3$, the result follows by inspection of the list of doubly transitive permutation groups (see chapter 12). (We remark that Kantor's original proof did not rely on the classification theorem.)

Finally, in the case $r=2$, we observe that $G \leqslant \mathrm{Aut}(T)$ for some tournament T, and G acts as a regular group on the set of edges of T. It follows by Theorem 1.3 that T must be a Paley tournament, hence n is a prime power and $\equiv -1 \pmod 4$. To see that in this case $\mathrm{KG}(n,r)$ is indeed a Cayley graph, let G be the group of affine transformations $x \mapsto ax+b$, $a,b,x \in \mathrm{GF}(n)$, a a square in $\mathrm{GF}(n)$. □

As this example shows, it is often not easy to decide whether or not a given vertex-transitive graph is a Cayley graph. If the number of vertices is a prime power, the following partial information is useful.

Theorem 3.4. (a) *If G is a transitive group of degree p^k, p prime, then the Sylow p-subgroups of G are transitive as well* (Wielandt 1964, p. 6).

(b) (Marušič 1985) *Every vertex-transitive (di)graph of order p^k, $k \leqslant 3$, is a Cayley (di)graph. Counterexamples exist for $k \geqslant 4$.*

Let $\mathcal{V}$ denote the set of those positive integers n for which there exists a connected vertex-transitive graph of order n which is not a Cayley graph. Considerable effort has gone into determining the set $\mathcal{V}$ (see the survey Praeger 1990). It is clear that all multiples of a member of $\mathcal{V}$ also belong to $\mathcal{V}$ (the complement of the disjoint union of copies of a non-Cayley vertex-transitive graph is again non-Cayley). So we need to know the minimal members of $\mathcal{V}$ only (w.r. to divisibility). It is not known whether or not such minimal members can have an arbitrarily large number of distinct prime divisors. It is conjectured that almost all vertex-transitive graphs of order n are Cayley graphs.

Cayley graphs are not edge transitive in general. (The triangular prism is an example.) In fact, their automorphism group often coincides with their group of definition (see the GRR problem in section 4.3). Here is a sufficient condition to guarantee added symmetry.

Proposition 3.5 (Frucht 1952). *If a group automorphism* $\alpha \in \mathrm{Aut}(G)$ *stabilizes the set* $S \subseteq G$ *then* $\alpha \in \mathrm{Aut}(\Gamma(G,S))$.

Corollary 3.6. (a) *If* S *is an orbit of some subgroup* H *of* $\mathrm{Aut}(G)$ *then* $\Gamma(G,S)$ *is edge-transitive.*

(b) *If, in addition,* $S = S^{-1}$, *then* $\Gamma(G,S)$ *is flag-transitive.*

(c) *An edge-transitive Cayley graph of an abelian group is flag-transitive.*

Note that the added condition in (b) is automatically satisfied if S consists of involutions (elements of order 2). Frucht (1952) employed this observation to construct a flag-regular graph of degree 3. Another application is the construction of 2-arc-transitive covering graphs (Theorem 1.5).

3.2. Symmetry and connectivity

The implications of vertex-transitivity to connectivity properties of graphs were discovered by Mader (1971a,b) and Watkins (1970). Their methods and results were generalized to directed graphs by Hamidoune (cf. Hamidoune 1981). We state the directed graph versions; undirected graphs are viewed as digraphs with edges oriented both ways. We note that a weakly connected finite vertex-transitive digraph is automatically *strongly connected* so we may omit the adjective. The *connectivity* $\kappa(X)$ of a strongly connected digraph $X \neq K_n$ is the minimum number of vertices whose deletion destroys strong connectivity. Edge-connectivity is defined similarly.

Theorem 3.7. *Let* X *be a finite connected vertex-transitive digraph of out-degree* d.

(a) *The connectivity of* X *is* $\geqslant \lceil (d+1)/2 \rceil$. *If* X *is undirected then* $\kappa(X) \geqslant \lceil 2(d+1)/3 \rceil$.

(b) *The edge-connectivity of* X *is* d.

(c) *If* X *is edge-transitive or vertex-primitive, then* $\kappa(X) = d$.

The bounds in part (a) are tight, as shown by the lexicographic product of a (directed or undirected) cycle of length $m \geqslant 4$ and K_r.

All these results are simple consequences of the theory of atoms, developed by Mader, Watkins, and Hamidoune in the same papers (cf. chapter 2, section 7.5). A *positive fragment* of a strongly connected digraph X is a subset $F \subset V(X)$ such that the set $X^+(F)$ of out-neighbors of F has cardinality $\kappa(X)$ and $F \cup X^+(F) \neq V(X)$ (so $X^+(F)$ is a minimum cutset). An *positive atom* is a positive fragment of minimum cardinality.

The key result of Mader, Watkins, and Hamidoune is that *if A is a positive atom and F is a positive fragment then either $A \subseteq F$ or $A \cap F = \emptyset$*. (For a simple proof, see Hamidoune 1981, Theorem 2.1.) In particular, the positive atoms are pairwise disjoint. Consequently, if X is vertex-transitive then the atoms form a system of imprimitivity. From this, the vertex-connectivity results readily follow. For the edge-connectivity result, edge-atoms are introduced and their disjointness proved. (Cf. also Lovász 1979a, chapter 12 for these and related results.)

Corollary 3.8 (Cauchy-Davenport). *Let $\emptyset \neq A, B \subset \mathbb{Z}_p$ (p a prime). Then* $|A + B| \geqslant \min\{p, |A| + |B| - 1\}$.

Proof. W.l.o.g., $0 \in B$. Apply part (c) of Theorem 3.7 to the vertex-primitive Cayley digraph $X = \Gamma(\mathbb{Z}_p, B \setminus \{0\})$. Conclude that if $A + B \neq \mathbb{Z}_p$ then $|X^+(A)| \geqslant \kappa(X) = |B| - 1$. Observe, on the other hand, that $X^+(A) = (A + B) \setminus A$. □

For this result and other connections with additive number theory, see Hamidoune (1990).

Minimal Cayley graphs do even better than guaranteed by part (a) of Theorem 3.7: *If X is a minimal Cayley graph of degree d then* $\kappa(X) = d$ (Godsil 1981a).

Infinite connected vertex-transitive graphs of arbitrarily large degree may have connectivity as small as 1, as the example of the regular tree of any degree demonstrates. Yet, analogous results exist. Let $\kappa_f(X)$ denote the smallest size of a subset C of the vertex set of a locally finite infinite graph X such that at least one of the connected components of $X \setminus C$ is finite. *If X is connected, vertex-transitive, and it has finite degree d, then* $\kappa_f(X) \geqslant \lceil 3(d + 1)/4 \rceil$ (Babai and Watkins 1980). Analogously to the finite case, the proof rests on the disjointness of atoms (finite sets of vertices with κ_f neighbors). We note that if the graph X has just one *end* (cf. section 3.7) then $\kappa(X) = \kappa_f(X)$.

3.3. Matchings, independent sets, long cycles

All graphs in this section are finite. The next question concerns *matchings*.

Theorem 3.9 (Little, Grant and Holton 1975). *Let X be a connected vertex-transitive graph on n vertices.*

(a) *If n is even then X has a perfect matching.*

(b) *If n is odd then X is matching critical.*

(c) (Lovász, Plummer) *If n is even then X is either bicritical (deletion of any pair of vertices leaves a perfect matching) or elementary bipartite (deletion of any pair of vertices of opposite color leaves a perfect matching).*

Proof. Let D be the set of those vertices of X which are left uncovered by at least one maximum matching. The Gallai–Edmonds structure theorem (see chapter 3, section 4.3) asserts that if D is not empty then it consists of matching critical components. But if X does not have a perfect matching then, by vertex-transitivity, D is the entire vertex set. This proves (a) and (b). For (c), see Lovász and Plummer (1986, Theorem 5.5.24). □

The following two observations on regular uniform hypergraphs come in handy in the analysis of various kinds of subsets of vertex-transitive graphs. (A hypergraph is *regular* of degree d if every vertex is contained in exactly d edges.)

Lemma 3.10 (Regular hypergraph counting lemma). *Let $\mathscr{E}$ and $\mathscr{F}$ be r-uniform and s-uniform regular hypergraphs, resp., on the same set of n vertices.*

(a) *Assume $|E_i \cap F_j| \geqslant d$ for every $E_i \in \mathscr{E}$, $F_j \in \mathscr{F}$. Then $rs \geqslant nd$.*

(b) *Assume $|E_i \cap F_j| \leqslant d$ for every $E_i \in \mathscr{E}$, $F_j \in \mathscr{F}$. Then $rs \leqslant nd$.*

If $\mathscr{E} = \mathscr{F}$ and $d \neq n$ then under condition (a) *we have $r^2 > nd$. On the other hand, if $\mathscr{E} = \mathscr{F}$, $d \neq n$, and $|E_i \cap E_j| \leqslant d$ for every $E_i, E_j \in \mathscr{E}$, $i \neq j$, then $r^2 < 2nd$.*

Proof. (a) Fix $E_i \in \mathscr{E}$. Count the number of pairs (x, j) such that $x \in E_i \cap F_j$. The result is $r \deg_{\mathscr{F}} \geqslant d|\mathscr{F}| = dn \deg_{\mathscr{F}} / s$. (b) This part, as well as the $\mathscr{E} = \mathscr{F}$ variants, follows analogously. □

As a corollary, we have a tradeoff between $\alpha(X)$, the maximum size of independent sets, and $\omega(X)$, the maximum size of cliques of the graph X, for vertex-transitive graphs. For a generalization in the context of the Shannon capacity of graphs, see Lovász (1979b) (cf. chapter 31, section 6).

Corollary 3.11 (L. Lovász and R. M. Wilson). *If X is a vertex-transitive graph then $\alpha(X)\omega(X) \leqslant n$.*

Proof. Indeed, let $\mathscr{E}$ and $\mathscr{F}$ be the hypergraphs consisting of the independent sets and cliques, resp., of maximum size. Each of these two hypergraphs is uniform by definition; they are regular because X is vertex-transitive. Since a clique and an independent set share at most one vertex, the result follows from part (b) of Lemma 3.10. □

We note that Delsarte (1973) proves the same conclusion under the condition that X is the union of classes in an association scheme (cf. chapter 15), in particular, if X is strongly regular. This condition does not imply the presence of any automorphisms, nor is it a consequence of vertex-transitivity.

A related observation concerns the connection between the chromatic number $\chi(G)$ and the independence number $\alpha(G)$. Clearly, $\chi(G) \geqslant n/\alpha(G)$ for all graphs. For vertex-transitive graphs, this inequality is nearly tight, as pointed out to us by M. Szegedy.

Proposition 3.12. *If G is a vertex-transitive graph then*

$$n/\alpha(G) \leqslant \chi(G) \leqslant n(1 + \ln n)/\alpha(G).$$

Proof (*of the rightmost inequality*). Let A be an independent set of size $\alpha = \alpha(G)$; then the probability that $m = \lceil \ln n \rceil$ random translates of A (by automorphisms) do not cover $V(G)$ is less than $n \cdot (1 - \alpha/n)^m < n \cdot \mathrm{e}^{-\alpha m/n} \leqslant 1$. □

Another corollary to Lemma 3.10, of interest to the theory of computing, concerns *Boolean functions* $f: 2^X \to \{0, 1\}$. Here, X is a set of n Boolean variables $x_1, \ldots, x_n$, and 2^X represents the set of all possible truth value assignments to X. A partial truth value assignment $y: Y \to \{0, 1\}$ $(Y \subseteq X)$ is said to *force* f to 0 if $f(x) = 0$ whenever x is an extension of y. We call such a y a *0-certificate* for f; its *size* is $|Y|$, the cardinality of the domain of y. We define *1-certificates* analogously. For every $x \in 2^X$ there exists a smallest restriction y of x which is an $f(x)$-certificate; let $m(f; x)$ denote its size, and let $n_i(f) = \max_x m(f; x)$ where the maximum is taken over all x with $f(x) = i$ $(i = 0, 1)$. Let $N(f) = \max\{n_0(f), n_1(f)\}$. The quantity $N(f)$ is called the *nondeterministic decision-tree complexity* of f. (This is a lower bound on the *deterministic* decision-tree complexity discussed in chapter 34, section 4.4. Incidentally, the "evasiveness" problems considered there relate symmetry to complexity in a remarkable way.)

Automorphisms of f are those permutations of X which leave f invariant.

Corollary 3.13. *If f is a non-constant Boolean function on n variables with transitive automorphism group then $n_0(f) n_1(f) \geqslant n$. Consequently, $N(f) \geqslant \sqrt{n}$.*

Proof. The domains of a 0-certificate and a 1-certificate must intersect. One can thus apply Lemma 3.10 (a) to the hypergraphs formed by an orbit of each kind of domain. □

Our next subject is *long paths* and *cycles*. Only four connected vertex-transitive graphs without Hamilton cycles are known (assuming the number of vertices is $n \geqslant 3$). Each of them is trivalent; and the first two are *3-arc-transitive* (cf. section 5.1): the Petersen graph (10 vertices), and the Coxeter graph (28 vertices) (see figures 8, 9 in chapter 1, section 4). (The automorphism group of the latter is PGL(2, 7), see Wong (1967); cf. Biggs 1973). The other two are obtained from these by replacing each vertex by a triangle (30 and 84 vertices, resp.). Each of these four graphs possesses a Hamilton path and none of them is a Cayley graph. A conjecture of Lovász (in 1969) *not shared* by this author holds that all connected vertex-transitive graphs have Hamilton paths. The problem as to whether all Cayley graphs $(n \geqslant 3)$ have Hamilton cycles appears first to have been stated by Rapaport-Strasser (1959). In my view these beliefs only reflect that Hamiltonicity obstacles are not well understood; and indeed, vertex-transitive graphs may provide a testing ground for the power of such obstacles. We conjecture that for some $c > 0$, there exist infinitely many connected vertex-transitive graphs (even Cayley graphs) without cycles of length $\geqslant (1 - c)n$.

We mention a useful Hamiltonicity obstacle. A graph is *tough* if, after deletion of any k of its vertices, the remaining graph has $\leqslant k$ connected components. Obviously, any Hamiltonian graph is tough; being non-tough is a Hamiltonicity obstacle. This obstacle breaks down for vertex-transitive graphs: *every connected* *vertex-transitive graph is tough.* Indeed, by Theorem 3.7(b), a d-regular connected vertex-transitive graph has edge-connectivity d, a circumstance that immediately implies toughness.

At any rate, the Hamiltonicity conjectures have been confirmed in a number of cases. One notable Hamilton cycle was even patented in 1953: the one constructed by F. Gray for the minimal Cayley graphs of the elementary abelian groups of order 2^d (the d-cube). A large number of papers has since referred to Hamilton cycles in Cayley graphs as "generalized Gray codes". (See the references in Conway et al. 1989.)

In the subsequent statements, every graph has $n \geqslant 3$ vertices.

It is easy to see that every Cayley graph of a finite abelian group is Hamiltonian (J. Pelikán, see Lovász 1979a, Ex. 12.17). Marušič, Witte, Keating, Dürnberger, and others succeeded in significantly relaxing the condition of commutativity. We refer to the survey by Witte and Gallian (1984) for details. One of the weakest known *sufficient conditions* for all Cayley-graphs of G to be Hamiltonian is that *the commutator subgroup of G is cyclic of prime power order* (Keating and Witte 1985; cf. Dürnberger 1985). Witte (1986) proved that *all Cayley digraphs of a p-group have a Hamilton cycle.*

So far, no non-solvable group has been shown to have this property. Even the following, less ambitious problem is open: does every finite group have a minimal Cayley graph with a Hamilton cycle?

For several reasons (including, as a curiosity, *campanology*, the study of bell ringing sequences, see White 1985), special classes of Cayley graphs of the symmetric groups are of interest.

Theorem 3.14 (Kompel'macher and Liskovets 1975). *Let T be any connected system of transpositions of n elements. Then the Cayley graph $\Gamma(S_n, T)$ is Hamiltonian.*

The case of adjacent transpositions (see Johnson 1963) was recently generalized to all finite reflection groups (groups of affine transformations of $\mathbb{R}^n$, generated by a set of reflections) (Conway, Sloane and Wilks 1989).

The situation for Cayley *digraphs* is more complicated. Rankin (1948) determined when a Cayley digraph $\Gamma(G,S)$ of a finite abelian group G is Hamiltonian provided $|S| = 2$ and gave examples of Cayley digraphs of the alternating groups A_6 and A_7 without Hamilton cycles. J. Milnor gave a class of solvable groups with two generators such that the difference between the order of the group and the longest directed paths in the resulting Cayley digraphs is arbitrarily large (see Witte and Gallian 1984).

Much less is known about vertex-transitive graphs of given order. Trivially, every connected vertex-transitive graph of prime order p is a Cayley graph and is therefore Hamiltonian. Marušič's result (Theorem 3.4(b)) extends this to orders p^2 and p^3.

Theorem 3.15. (a) *Every connected vertex-transitive graph of order n is Hamiltonian if n has one of the following forms: p, $2p$, with the exception of the Petersen graph* (Alspach 1979); *$3p$, p^2, p^3, $2p^2$* (Marušič 1985, 1987);

(b) *Every connected vertex-transitive graph of order* $4p$ *and* $5p$ *has a Hamilton path* (Marušič and Parsons 1983).

The only general lower bound on the length of the longest cycles and paths of vertex-transitive graphs is the following. (Nothing better is known for Cayley graphs either.)

Proposition 3.16 (Babai 1979b). *If* X *is a connected vertex-transitive graph on* $n \geqslant 5$ *vertices then* X *has a cycle of length* $> 2\sqrt{n}$.

We note that 3-connected trivalent graphs have cycles of length $\geqslant n^{0.69}$ (Jackson 1986) but need not have cycles longer than $n^{0.96}$ (Bondy and Simonovits 1980).

The proof of Proposition 3.16 is based on the following observation: *If* X *is a 3-connected regular graph of order* $n \geqslant 4$ *then every pair of longest cycles intersects in* $\geqslant 4$ *vertices.* Now, since every connected vertex-transitive graph of degree $\geqslant 3$ is 3-connected (Theorem 3.7(a)), an application of the Regular hypergraph counting lemma 3.10 to the vertex sets of the longest cycles completes the proof of Proposition 3.16. □

3.4. *Subgraphs, chromatic number*

Every graph Y with n vertices is an induced subgraph of some Cayley graph X of any given group of order $\geqslant cn^2$ (Babai 1978b, Babai and Sós 1985, Godsil and Imrich 1987). Every Y can be embedded into a Cayley graph of order 2^n such that all automorphisms of Y extend to automorphisms of X. The following more general extension theorem holds.

Theorem 3.17 (Hrushovski 1992). *Given a finite graph* Y *and a family* $\mathcal{F}$ *of isomorphisms between pairs of induced subgraphs of* Y, *there exists a finite graph* X *containing* Y *as an induced subgraph, such that all elements of* $\mathcal{F}$ *extend to automorphisms of* X.

Here, X may be required to be flag-transitive. This result has applications to the structure theory of the automorphism group of the Rado graph (countable "random graph", cf. section 5.3). It follows that "almost all" automorphisms of that graph are conjugates ("almost all" in the sense "comeager" (complement of a set of first Baire category; cf. Oxtoby 1980): there exists a *comeager* conjugacy class; cf. Truss 1992).

Not all graphs are subgraphs of *minimal* Cayley graphs. Let $X = \Gamma(G, S)$ be a minimal or semiminimal Cayley graph (cf. section 3.1) of the (finite or infinite) group G. Such graphs admit a coloring of the edges with the following properties: (a) every vertex has degree $\leqslant 2$ in each color; (b) at least one of the colors occurring in a cycle occurs at least twice on that cycle. (In the minimal case, each color occurring in a cycle occurs at least twice.)

These properties put constraints on the possible subgraphs. In particular, if X is a minimal Cayley graph then it contains no K_4^- (K_4 minus an edge), and no $K_{2,3}$.

If X is semiminimal, it contains no $K_{5,17}$ (Babai 1978a). In both cases it follows that the chromatic number of X is at most countably infinite, according to the following result of Erdős and Hajnal (see chapter 42, Theorem 6.3.): *If a graph has uncountably infinite chromatic number then it contains $K_{m,\aleph_1}$ for every positive integer m.*

It is an *open problem* whether or not the chromatic number of finite minimal Cayley graphs is bounded. We conjecture it is not. A related stronger conjecture is that for every $\varepsilon > 0$ there exist minimal Cayley graphs X such that $\alpha(X) \leqslant \varepsilon|V(X)|$ where $\alpha(X)$ denotes the size of the largest independent set of X.

A strong consequence of constraints (a) and (b) above was deduced by Spencer.

Theorem 3.18 (Spencer 1983). *For every $g \geqslant 3$ there exists a finite graph Y of girth g such that Y is not a subgraph of any (semi)minimal Cayley graph.*

The proof uses the probabilistic method and does not provide explicit graphs Y. It is not known whether or not such excluded subgraphs of girth 5 and degree 3 exist, even for minimal Cayley graphs. (The Petersen graph is a subgraph of a minimal Cayley graph of a group of order 20.)

Every finite group has a Cayley graph of chromatic number $\leqslant 4$. (This is a consequence of the fact that every finite simple group is generated by $\leqslant 2$ elements.) It is an open question whether or not every infinite group has a Cayley graph of finite chromatic number.

3.5. Neighborhoods, clumps, Gallai–Aschbacher decomposition

In this section we highlight a graph theoretic result that has played a role in the *classification theory of finite simple groups.*

We shall (in this section) consider finite graphs as well as locally finite infinite graphs with uniformly bounded degrees. X will always denote a graph with vertex set V and complement $\overline{X}$; the set of neighbors of $x \in V$ is denoted by $X(v)$. The subgraph induced by $X(v)$ is the *link* at v. We say that X has *constant link Y* if all of its links are isomorphic to Y (a finite graph by the convention above). All vertex-transitive graphs have constant link, and many others, including all triangle-free regular graphs and their line graphs. A finite graph Y is a *link graph* if there exists a graph X with constant link Y. If such a finite X exists then Y is a link of finite type.

Many classes of link-graphs as well as non-link graphs have been found (see Hell 1978, Blass et al. 1980). However, Bulitko (1972) asserts that the problem whether or not a given finite graph is a link graph is undecidable. It is shown in Bulitko (1972) and Brown and Connelly (1975) that there exist graphs which are links of infinite vertex-transitive graphs but do not occur as links in finite constant-link graphs. By counting certain triangles, Blass et al. (1980) show that *if L is the link in a finite vertex-transitive graph and i is an odd number then the number of vertices of degree i in L is even.* This is not true for all link graphs of finite type (Brown and Connelly 1975 provides an infinity of examples), nor does it hold for links of infinite vertex-transitive graphs. Hell (1978) observes that if a (finite or infinite) vertex-transitive graph X has an *asymmetric* link then X is a Cayley graph (in fact a GRR, cf. section 4.3). He also shows that the link of a Cayley graph has an even number of vertices of degree one.

The following fairly general result is implicit in Aschbacher (1976).

Theorem 3.19. *Assume both X and $\overline{X}$ are connected. Then at least one of the links of X, say Y, has the property that $\overline{Y}$ has a unique largest connected component.*

A stronger result holds for vertex-primitive graphs.

Theorem 3.20 (Aschbacher, Fischer). *Let X be a vertex-primitive graph other than the complete graph. Let Y be the graph induced by the neighborhood of a vertex in X. Then the complement of Y is connected.*

The proof of these theorems rests on a purely graph theoretical result, part of which was discovered by Gallai (1971) in the context of the *characterization of transitively orientable graphs.*

A subset $C \subseteq V$ is called a *clump* if for each $w \in V \setminus C$, if w has a neighbor in C then $X(w) \supseteq C$. The *trivial* clumps are V, $\emptyset$, and the singletons. A *proper* clump is a clump other than V. A *maximal* clump is a proper clump not properly contained in any other proper clump.

We begin with two easy observations. (a) If C, D are clumps and $C \cap D \neq \emptyset$ then $C \cup D$ is a clump. (b) If both X and its complement $\overline{X}$ are connected then V is not the union of two proper clumps. It is immediate from these that *maximal clumps are pairwise disjoint*, which proves part (i) of the following result.

Theorem 3.21 (Gallai–Aschbacher decomposition). *Assume both X and $\overline{X}$ are connected. Let $C_1, \ldots, C_m$ be the maximal clumps of X. Then:*

(i) (Gallai 1971, Aschbacher 1976) $(C_1, \ldots, C_m)$ *form a partition of V.*

(ii) (Aschbacher 1976) *Let N_i be the set of common neighbors of C_i. (By definition, $C_i \cap N_i = \emptyset$.) Then there exists i such that the subgraph induced by N_i in the complement of X is connected.*

To see how Theorem 3.20 follows, we observe that the maximal clumps form a system of imprimitivity for Aut(X); therefore if X is vertex-primitive then each C_i is a singleton.

The proof of assertion (ii) is nontrivial. For $v \in V$, consider the components of the subgraph of $\overline{X}$ induced by $X(v)$. Let M be a maximal such component (considering all $v \in V$). By definition, M induces a connected subgraph of $\overline{X}$. Let C be the set of common neighbors of M. One can prove that C is a maximal clump, and M is the set of common neighbors of C. This completes the proof of Theorem 3.21; and together with (i) above we also see that the choice of M among the components of $X(v)$ in $\overline{X}$ must be unique (since any other component is a subset of C, the unique maximal clump containing v). □

Gallai (1971) gives the following equivalent definition of the above decomposition. Let us say that two edges are equivalent if they together form an induced path of length 2. Take the transitive closure of this relation to obtain the *Gallai equivalence.* If both X and $\overline{X}$ are connected, then there will be a unique Gallai class of edges which spans the entire X. The components of the complement of this class can be grouped together in a unique way to produce the maximal clumps; two such components will belong to the same class if they have the same neighborhood in X.

The role of Theorem 3.20 in the *classification of finite simple groups* is explained by Aschbacher (1976). He shows how Fischer's (1971) celebrated "3-transpositions theorem" follows from it; in fact, the result arose from one of Fischer's lemmas. A set of 3-*transpositions* is a set S of elements of order 2 in a group G such that for any pair $g, h \in S$, the order of gh is $\leqslant 3$. Fischer characterized those almost simple groups which are generated by a conjugacy class of 3-transpositions. These include all the symmetric groups, certain classical (symplectic, orthogonal, unitary) groups, plus three sporadic groups discovered in the process (named $M(22)$, $M(23)$, and $M(24)$). ($M(24)$ is not simple; like the symmetric groups, it has a simple subgroup of index 2. Cf. Aschbacher 1980.)

Fischer's central result was that the action of G by conjugation on S is a rank 3 permutation group. This is derived from considering the vertex-transitive graph with vertex set S, joining two elements if they commute. G is shown to act as a primitive group on this graph; and Theorem 3.20 is invoked. □

Godsil (1980b) considers the link L of X together with the link L^* of $\overline{X}$, the *dual link.* He gives the following remarkable characterization: *if X is finite, vertex-transitive, both the link L and the dual link L^* are disconnected but at least one of them has no isolated vertices, then $X \cong L(K_{3,3})$.* He also characterizes the case when both L and L^* have isolated vertices. In this latter case, $\mathrm{Aut}(X)$ always has an element of the form $(12)(34)$. These results are central to his solution of the GRR problem (cf. section 4.3).

3.6. Rate of growth

Note. Throughout this section, X will denote an *infinite, connected, locally finite* graph. (A graph is locally finite if all vertices have finite degree.)

Certain properties of groups are best expressed in graph theory language. A foremost example is the *growth rate* of finitely generated infinite groups.

For a graph X, let $B(n, x)$ denote the *ball of radius n* about the vertex x, i.e. set of vertices at distance $\leqslant n$ from x. For a vertex-transitive graph, set $f(n) = |B(n, x)|$. This function has a property resembling log-concavity.

Proposition 3.22 (Gromov 1981). *If X is vertex-transitive, then $f(n)f(5n) \leqslant (f(4n))^2$.*

Proof. Let Y be a maximal system of vertices in $B(3n, x)$ pairwise at distance $\geqslant 2n + 1$. Now the disjoint balls $B(n, y)$: $y \in Y$ are contained in $B(4n, x)$, hence $|Y| f(n) \leqslant f(4n)$. On the other hand, the balls $B(2n, y)$: $y \in Y$ cover $B(3n, x)$, and therefore the balls $B(4n, y)$: $y \in Y$ cover $B(5n, x)$. This implies $f(5n) \leqslant |Y| f(4n)$, hence the result. □

X is said to have *growth rate* $g(n)$ if $g(c_1 n) \leqslant f(n) \leqslant g(c_2 n)$ for positive constants c_1, c_2 and every sufficiently large n. Thus, the growth rate is an equivalence class of functions rather than a function. There is a natural partial order on the equivalence classes; when comparing growth rates, we shall always mean comparison of their equivalence classes.

X is said to have *polynomial growth rate* if its growth rate is bounded by n^c for some constant c; its growth rate is *exponential* if it is bounded from below by c^n for some constant $c > 1$.

For a finitely generated infinite group G, the *growth rate of* G is defined as the growth rate of the Cayley graph $\Gamma(G, S)$ for some finite set S of generators of G. It is easy to see that the growth rate does not depend on the particular choice of S; a change in the generators will only affect the constants c_1 and c_2.

Finitely generated abelian groups have polynomial growth rates, non-cyclic free groups have exponential growth rates. The following are easy to prove.

Proposition 3.23. (a) *If H is a subgroup of G, then the growth rate of G is greater than or equal to the growth rate of H.*

(b) *If $|G{:}H|$ is finite then G and H have the same growth rates.*

(c) (Gromov 1981) *If H is finitely generated and $|G{:}H|$ is infinite then $f_G(n) \geqslant n f_H(n)$, where f_G and f_H are the growth functions of the respective groups under appropriately chosen sets of generators.*

(d) (Milnor 1968a, Wolf 1968) *Finitely generated nilpotent groups have polynomial growth rates.*

The Bass–Wolf formula gives the exact growth rates of nilpotent groups. Let G be a finitely generated infinite nilpotent group and let $G = G_1 > G_2 > \cdots > G_m = 1$ be its descending central series. Let d_i be the torsion-free rank of the abelian group G_i/G_{i+1}.

Theorem 3.24 (Bass 1972, Wolf 1968). *The rate of growth of the nilpotent group G is n^d where $d = \sum i d_i$.*

The following very deep result settles a problem raised by Milnor (1968a).

Theorem 3.25 (Gromov 1981). *A group has polynomial growth rate if and only if it is virtually nilpotent, i.e. it has a nilpotent subgroup of finite index.*

Two important particular cases of this result were established earlier; they are ingredients in Gromov's proof.

Theorem 3.26 (Milnor 1968a, Wolf 1968, Tits 1972). (a) *A finitely generated solvable group G has exponential growth unless G is virtually nilpotent.*

(b) *A finitely generated subgroup G of a connected Lie group has exponential growth unless it is virtually nilpotent.*

In fact, Tits proves the following stronger statement.

Theorem 3.27 (Tits 1972). *If L is a Lie group with finitely many components and G is a finitely generated subgroup of L then either:*

(a) *G contains a free group of rank 2 and has therefore exponential growth; or*

(b) *G is virtually solvable. In this case it has exponential growth rate unless it is virtually nilpotent.*

We give a very rough sketch of the proof of Gromov's theorem 3.25. Let G be a finitely generated group of polynomial growth. Fix a finite set S of generators. Select a sequence $r_i \to \infty$ of integers. Consider the sequence of metric spaces Γ_i on the set G with distance $d_i(x,y) = (1/r_i)\text{dist}(x,y)$ where "dist" is the distance in the Cayley graph $\Gamma(G,S)$. The sequence r_i is chosen so as to ensure a fairly regular behavior of the sequence $f(2^j r_i)$, $i = -j, \ldots, j$. This is accomplished with the aid of Proposition 3.22 and using the assumption of polynomial growth. The sequence Γ_i is then nice enough to have a subsequence that converges in an appropriate sense to a metric space Y. Elementary considerations show that Y is locally compact, connected, and locally connected. Moreover, each ball in Y is path-connected. The isometry group L of Y is transitive on Y. The choice of the r_i ensures that the Hausdorff dimension of Y is finite. A celebrated theorem of Montgomery and Zippin (1955) now implies that under these conditions, L is a Lie group with a finite number of components. Now a fairly involved argument using the quoted result of Tits (Theorem 3.27(b)) completes the proof. □

Other ingredients of this last part of the proof are the Milnor–Wolf theorem (Theorem 3.26(a)) and the following theorem of Jordan (cf. Raghunathan 1972).

Theorem 3.28 (Jordan 1895). *If L is a Lie group with a finite number of components then there exists a number q such that every finite subgroup of L contains an abelian subgroup of index at most q.*

An appendix to Gromov's paper contains a relatively simple proof of the subcase of the Milnor–Wolf theorem used in Gromov's proof.

Milnor (1968a) raised the question whether groups with "intermediate growth rates" (neither polynomial, nor exponential) exist. The positive answer was given by Grigorchuk.

Theorem 3.29 (Grigorchuk 1983). *There exist 2-generated torsion groups with growth rates between 2^{n^α} and 2^{n^β} where $\alpha = \frac{1}{2} - \varepsilon$ for any $\varepsilon > 0$ and $\beta = \log_{32} 31$.*

Vertex-transitive graphs with polynomial growth rates were characterized by V.I. Trofimov.

Theorem 3.30 (Trofimov 1985). *Let X be vertex-transitive. The following are equivalent.*

(a) *X has polynomial growth.*

(b) *The vertex set V under the action of $\text{Aut}(X)$ admits a system of imprimitivity σ with finite equivalence classes such that $\text{Aut}(X/\sigma)$ is finitely generated, virtually nilpotent, and the stabilizer of any vertex of X/σ in $\text{Aut}(X/\sigma)$ is finite.*

Here X/σ is the homomorphic image of X under the vertex map $V(X) \to V(X)/\sigma$; hence two equivalence classes are adjacent if they have at least one pair of adjacent representatives.

Related topics are surveyed in Trofimov (1992).

We should mention that these questions were originally motivated by connections between the curvature of a Riemannian manifold and the growth rate of its fundamental group (Milnor 1968b).

3.7. Ends

Note. Throughout this section, X will denote an *infinite, connected, locally finite* graph. (A graph is locally finite if all its vertices have finite degree.)

Ends are another important graphic notion for finitely generated infinite groups. (For a detailed account, see Cohen 1972.)

The set of *ends* of a connected, locally connected, locally compact Hausdorff space X is defined as the inverse limit of the directed family of the set of components of $X \setminus C$ for all compact subsets C (Hopf 1944). The analogous concept for connected graphs was developed by Halin (1964).

Ends of a (connected, infinite, locally finite) graph X can be defined analogously as the inverse limit of the sets of infinite components obtained by deleting finite subsets C of the edge set of the graph X. The ends can also be defined as equivalence classes of one-way infinite paths: two such paths are equivalent if the deletion of no finite set of edges separates their infinite components.

The *ends* of a finitely generated group are defined as the ends of its Cayley graphs. Different choices of finite sets of generators result in topologically equivalent sets of ends. Stallings (1971) contains important results on ends of groups.

Proposition 3.31 (Hopf 1944). *If X is vertex-transitive then it has* 1 *or* 2 *or infinitely many ends. In particular, the same holds for finitely generated infinite groups. Consequently, if X has more than* 2 *ends then it has exponential growth rate.*

A vertex-transitive graph has *two ends* if and only if it has *linear growth rate.* Groups with two ends have been fully characterized.

Theorem 3.32 (Freudenthal 1945). *A finitely generated infinite group G has two ends if and only if G has a finite normal subgroup N such that the quotient group G/N is either cyclic ($\mathbb{Z}$) or dihedral (the free product $\mathbb{Z}_2 * \mathbb{Z}_2$).*

Groups with infinitely many ends have also been characterized. We note that they have exponential growth rates; the converse is false. Let A be a group, F a subgroup, and $\varphi: F \to A$ an injection. The HNN-extension $G = (A, F, \varphi)$ is generated by A and an additional element x subject to the relations $x^{-1}fx = \varphi(f)$ $(f \in F)$.

Theorem 3.33 (Stallings 1971). *A finitely generated group G has infinitely many ends if and only if G is*

(a) *either a free product with amalgamated finite subgroup* $G = G_1 *_F G_2$, *where* F *is a finite proper subgroup of each* G_i *and has index* $\geqslant 3$ *in at least one of them;*

(b) *or an HNN-extension* $G = (A, F, \varphi)$, *where* F *is a finite proper subgroup of* A.

These cases are closely related to group actions on trees. A theory of such actions was developed by Bass and by Serre (1980). We quote two special cases.

The group G is said to act *without inversions* on a graph if no element of G inverts any edges. In other words, G preserves an orientation of the graph.

Theorem 3.34 (Serre 1980, chapter 4). (a) *Let* G *act edge-transitively but not vertex-transitively on a tree* T. *Let* P, Q *be two adjacent vertices of* X. *Then* G *is the free product of the stabilizers of* P *and* Q *amalgamated at their intersection.*

(b) *Every amalgam of two groups acts on a tree in this way.*

Theorem 3.35 (Serre 1980, chapter 5.4). *Let* G *act edge-transitively and vertex-transitively but without inversions on a tree* T. *Then* G *is an HNN-extension of the stabilizer of a vertex. Every HNN-extension acts on a tree in this way.*

For the general structure theorem, see section 3.11.

M.J. Dunwoody used Theorems 3.34 and 3.35 to give the following remarkable generalization of the Stallings characterization theorem (Theorem 3.33).

Theorem 3.36 (Dunwoody 1982). *Let* G *be a group acting on a connected graph* X *with* $\geqslant 2$ *ends. Then* G *is either an amalgam* $G = A *_C B$ *or an HNN-extension of a group* C, *where in each case* C *contains the stabilizer of two adjacent vertices as a subgroup of finite index.*

The proof is based on the construction of a tree T on which G acts without inversions and so that the quotient graph has a single edge. A key tool for the construction of T is the following surprisingly strong statement on the existence of cuts of very special kind.

Theorem 3.37 (Dunwoody 1982). *Let* $X = (V, E)$ *be a connected graph with* $\geqslant 2$ *ends. Then there exists a nonempty proper subset* $A \subset V$ *such that*

(a) *the set of edges between* A *and* $V \setminus A$ *is finite;*

(b) *for any* $g \in G$, *either* A *or* $V \setminus A$ *is included in either* A^g *or in* $(V \setminus A)^g$.

In this result, the graph X is not required to be locally finite.

3.8. Isoperimetry, random walks, diameter

The *boundary* of a subset U of the vertex set V of the graph X is the set ∂U of vertices in $V \setminus U$, adjacent to at least one vertex in U. The *isoperimetric ratio* of a set $W \subset V$ is defined as $\varepsilon(W) = |\partial(W)|/|W|$. We say that W is *ε-expanding* if $\varepsilon(U) \geqslant \varepsilon$ for every $U \subseteq W$ $(U \neq \emptyset)$. We call X an *ε-expander* if every subset $W \subset V$ with $1 \leqslant |W| \leqslant |V|/2$ is ε-expanding. A "family of linear expanders" is an infinite sequence of graphs of bounded degree which are ε-expanding for some fixed $\varepsilon > 0$. ("Linear" refers to the $O(V)$ bound on the number of edges.) Expanders are treated in detail in chapter 32. Some Cayley graphs of linear groups turn out to be particularly strong expanders (Lubotzky, Phillips and Sarnak 1988, Margulis 1988).

Here we shall focus on more modest expansion properties shared by *all vertex-transitive graphs*. The generality of the results is important in applications to the analysis of algorithms in groups (cf. Babai 1991a).

It is easy to see that that the diameter of an ε-expander on n vertices can be bounded as

$$\operatorname{diam}(X) < \ln n / \ln(1+\varepsilon). \tag{4}$$

For $\varepsilon \leqslant \frac{1}{2}$, we infer $\varepsilon < (\frac{4}{3})(\ln n/\operatorname{diam}(X))$. It is remarkable that for vertex-transitive graphs, this inequality is tight apart from an $\ln n$ factor.

Theorem 3.38 (Aldous 1987, Babai 1991b, Babai and Szegedy 1992). *If X is a vertex-transitive graph of diameter Δ then it is a $2/(2\Delta+1)$-expander.*

Aldous' proof (for Cayley graphs) is based on the following observation, due to Erdős and Rényi (1965) (cf. Babai and Erdős 1982), which we quote in a slightly generalized form (Babai and Sós 1985, Cooperman, Finkelstein and Sarawagi 1990).

Proposition 3.39. *Let G be a transitive group acting on a set V, $|V| = n$. Let $A \subseteq V$. Then*

$$(1/|G|) \sum_{g \in G} |A \cap A^g| = |A|^2/n. \tag{5}$$

(A set and its translates are "independent on average".)

It follows by greedy selection that G has a transitive subgroup generated by at most $\log_2 n + \log_2 \ln n + 1$ elements (Babai and Sós 1985), and a set of $O(\log n)$ random elements are likely to generate a transitive subgroup (Cooperman et al. 1990), a fact with implications to efficient manipulation of permutation groups (cf. Babai, Cooperman, Finkelstein and Seress 1991).

Let now $X = \Gamma(G, S)$ where $S = S^{-1}$ generates G and assume $\operatorname{diam}(X) = \Delta$. If $|A| \leqslant |G|/2$ then by (5) there exists $g \in G$ such that $|A \setminus gA| \geqslant |A|/2$. Aldous observes that $g = h_1 \cdots h_k$ for some $k \leqslant \Delta$ from which one concludes that $|h_i A \setminus A| \geqslant |A|/(2\Delta)$ for some some $h_i \in S$, proving a $1/(2\Delta)$ lower bound for Theorem 3.38. □

By Alon's (1986) theorem (see chapter 32, Theorem 3.2) it follows that for vertex-transitive graphs X of degree d and diameter Δ, the eigenvalue gap is $d - \lambda_2 > 1/(2\Delta+2)^2$, where λ_2 is the second largest eigenvalue of X.

This eigenvalue gap is significant in estimating the speed at which a *random walk* over X approaches the uniform distribution. Let us consider a *lazy* random walk on X, in which at every step we flip a coin; if it comes out heads, we do not move, else we move to an adjacent vertex, each neighbor having equal chance of being visited. (This trick eliminates potentially annoying negative eigenvalues from the matrix of transition probabilities.) A direct consequence of the foregoing considerations is the following rapid convergence (Aldous 1987, Babai, 1991b).

Corollary 3.40. *Let v_0, v_1 be vertices of a vertex-transitive graph of degree d and diameter Δ with n vertices. After ℓ steps, the lazy random walk, starting at v_0, will be at v_1 with probability $(1/n)(1 \pm \varepsilon)$, where*

$$\varepsilon \leqslant n\exp(-\ell/(8d \cdot (\Delta + 2)^2)). \tag{6}$$

In particular, if both d and Δ are bounded by $(\log n)^{O(1)}$, then so is the time it takes for the lazy random walk to arrive at a nearly uniformly distributed place.

For specific Cayley graphs (related, e.g., to card shuffling), different methods have been used to obtain strong estimates on the time it takes to reach near uniformity (Aldous 1983, Aldous and Diaconis 1987, Diaconis 1988).

While results of this kind necessarily require the graph to have small diameter (cf. (4)), vertex-transitive graphs with large diameter, including infinite graphs, also possess a similar *local expansion* property.

Theorem 3.41 (Local expansion, Babai 1991b, Babai and Szegedy 1992). *Let X be a connected (finite or infinite) vertex-transitive graph with vertex set V. Assume that the finite subset $U \subset V$ is within the ball of radius t about some vertex; and $|U| \leqslant |V|/2$. Then U is a $2/(2t+1)$-expanding set.*

When $X = \Gamma(G,S)$ is a Cayley graph, again a single generator is responsible: $|Ug \setminus U| \geqslant |U|/(4t)$ for some generator $g \in S$ (Babai 1991b).

This result is a tool in the rigorous analysis of efficient algorithms for permutation groups (Babai, Cooperman, Finkelstein and Seress 1991). A further consequence is that in vertex-transitive graphs, random walks *do not get stuck in a corner* for too long. In the theorem below, $X^k(v)$ denotes the ball of radius k about vertex v, and we consider how soon a random walk, starting at v, may be expected to be outside this ball.

Theorem 3.42 (Babai 1991b). *Let v be the start vertex of a random walk over a connected vertex-transitive graph X of finite degree d. Assume $|X^{4k}(v)| \leqslant |V|/2$. Let $\ell \geqslant ck^2d \cdot \ln|X^{4k}(v)|$. Then with probability $\geqslant \frac{1}{16}$, at a random time chosen uniformly from $\{1, 2, \ldots, \ell\}$, the random walk will be outside $X^k(v)$. (c is an appropriate constant.)*

This result is at the heart of an algorithm which, given a set of generators of a finite group G, constructs *nearly uniformly generated random elements of G* in $O(|\log(G)|^5)$ group operations (Babai 1991b). Reducing the exponent 5 would be of great significance, since many algorithms in computational group theory rely on "randomly chosen" elements from the group (see, e.g., Neumann and Praeger 1992, Beals and Babai 1993). The fast heuristics currently used to select such elements have as yet resisted rigorous analysis (cf. Celler et al. 1995).

Random walks over locally finite *infinite* graphs such as the d-dimensional grid have been of great interest for their many applications which include approximations for partial differential equations and curvature of Riemannian manifolds (see Kesten 1959 and the references in Thomassen 1990, Markvorsen et al. 1994). One of the basic qualitative properties of such graphs is whether they are recurrent (random walks return to their start with probability 1) or transient (with positive probability they never return). A classical result of Pólya (in 1921) (see Feller 1968, vol. I, 14.7) asserts that $\mathbb{Z}^2$ is recurrent, while $\mathbb{Z}^3$ is transient. For connections of this theory with electrical currents, see Doyle and Snell (1984), Thomassen (1990). Expansion properties play a critical role in determining transience; if for some fixed $\varepsilon > 0$ we have $|\partial U| \geqslant |U|^{1/2+\varepsilon}$ for every finite $U \subset V$ then the graph is transient (Varopoulos 1985, 1991). This result is tight in the sense that $\varepsilon = 0$ would not suffice, as the plane grid $\mathbb{Z}^2$ shows.

For Cayley graphs of finitely generated groups, transience/recurrence does not depend on the choice of the set of generators. Transience is inherited by subgroups of finite index; recurrence is inherited by all subgroups.

Thomassen and Woess (1994) survey a large body of literature on related topics.

Our next subject is the *diameter* of Cayley graphs (cf. Babai et al. 1990 for more references). A regular graph of degree $r \geqslant 2$ and diameter d has at most

$$n \leqslant 1 + r + r(r-1) + \cdots + r(r-1)^{d-1} = 1 + r((r-1)^d - 1)/(r-2)$$

vertices, hence $d > \log_{r-1}(n/3)$. The construction of Cayley graphs of given degree and small diameter is motivated, among others, by interconnection network design for parallel computer architectures. Bounds on the diameter with respect to given generators are relevant for puzzles like Rubik's cube: in this case, the question is the diameter of a specific Cayley graph of a group of order 43 252 003 274 489 856 000 with respect to a set of 12 generators. (The diameter is known to be $\geqslant 19$ and a rigorous almost certain probabilistic proof exists that it is no more than 36 (Fiat et al. 1989).)

As noted above, expanders have diameter $O(\log n)$. For its simplicity and small diameter, interconnection network designers favor a Cayley graph which is not an expander: the *cube-connected cycles.* (Cf. Leighton 1992.) This is the Cayley graph of the group $\mathbb{Z}_2 \wr \mathbb{Z}_s$ (of order $n = s2^s$), with generators τ, ρ where τ is an involution from the first copy of $\mathbb{Z}_2$, and ρ is a rotation of order s, permuting the s copies of $\mathbb{Z}_2$. The vertices can be represented by $(0,1)$-strings of length s with one position marked. The neighbors of such a marked string are obtained by switching the marked symbol, or moving the mark left or right by one position, viewing the rightmost and leftmost positions adjacent. The graph has degree 3 and diameter $\lfloor 5s/2 \rfloor - 2$, whereas $\log_2 n = s + \log_2 s$.

We note that not all groups of order n with k generators admit Cayley graphs of degree $O(k)$ and diameter $O(\log n)$. Groups with a bounded number of generators and a nilpotent subgroup of bounded index and bounded class of nilpotence require diameter n^c by Proposition 3.23(b) and (d). Using commutator collection, Annexstein and Baumslag obtained the following explicit value.

Theorem 3.43 (Annexstein and Baumslag 1989). *Let G be a group of order n with a nilpotent subgroup of index t and class ℓ. If S is a set of k generators of G then the diameter of $\Gamma(G,S)$ is at least $(n/t)^c$, where*

$$c = (kt\ell)^{-\ell}/2.$$

(For abelian subgroups, $\ell = 1$.)

Theorem 3.44 (Babai, Kantor and Lubotzky 1989). *Every nonabelian finite simple group G of order n has a set S of at most 7 generators such that the diameter of $\Gamma(G,S)$ is $\leqslant C \log n$ for some absolute constant.*

The Cayley graphs constructed in the proof are unlikely to be expanders; it is not known whether an expander family of bounded degree Cayley graphs of the alternating groups exists, for instance. They have the advantage, however, that, given an element of G in the natural (matrix) representation of G, there is an efficient algorithm (polynomial in $\log n$) to solve this "generalized Rubik's cube" puzzle, i.e. to compute a path of length $O(\log n)$ to the identity. (The explicit expanders mentioned in chapter 32 give no clue, how to find such a short path.) As an illustration, we describe the solution for the case $G = \mathrm{SL}(2,p)$. With the generators

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

we obtain expanders but no explicit routing. Instead, we choose the generators

$$C = \begin{pmatrix} 2 & 0 \\ 0 & \frac{1}{2} \end{pmatrix}, \quad D = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

and A. It is easy to see that A and C rapidly generate all strict upper triangular matrices because

$$C^{-1} \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix} C = \begin{pmatrix} 1 & 2k \\ 0 & 1 \end{pmatrix}.$$

Conjugating by D we obtain transposes.

Kantor (1992) proves that for $n \geqslant 10$, the groups $\mathrm{PSL}(n,q)$ have trivalent Cayley graphs of logarithmic diameter $O(n^2 \log q)$.

A particularly elegant construction of a Cayley graph of the symmetric group S_n, having diameter $\leqslant 6.75 n \log_2 n$, was given by Quisquater (1986) (cf. Babai, Hetyei, Kantor, Lubotzky and Seress 1990).

If we admit a logarithmic number of generators, the situation becomes favorable for every group. The following result is a consequence of Proposition 3.39 (cf. Babai and Erdős 1982 for a short proof).

Theorem 3.45 (Erdős and Rényi 1965). *Given a group G of order n, there exists a set of $k \leqslant \log_2 n + \log_2 \ln n + 1$ elements $S = \{g_1, \ldots, g_k \in G\}$ such that every $x \in G$ is representable in the form*

$$x = g_1^{\varepsilon_1} g_2^{\varepsilon_2} \cdots g_k^{\varepsilon_k}, \quad \text{where } \varepsilon_i \in \{0,1\}.$$

In particular, the diameter of $\Gamma(G,S)$ is $\leqslant k$.

In estimating the diameter of a Cayley graph, one faces much greater difficulties if the generators are prescribed. We conjecture, that for every finite simple group G and every set S of generators, the diameter of $\Gamma(G,S)$ is at most $(\log|G|)^c$ for some absolute constant c. Even in the case of alternating, or, equivalently, symmetric groups, this has only been verified in very special cases. For permutation groups, the following are known.

Theorem 3.46. *Let $G \leqslant S_n$ be generated by the set S. Then the diameter of $\Gamma(G,S)$ is not greater than:*

(a) $c_k n^2$, *if all members of S are cycles of lengths $\leqslant k$* (Driscoll and Furst 1987);

(b) cn^{2k}, *if all members of S have degree $\leqslant k$* (McKenzie 1984);

(c) $\exp(\sqrt{n \ln n}(1 + o(1)))$, *if no assumption on S is made* (Babai and Seress 1988).

The bound in (c) is asymptotically tight, as shown by the cyclic group generated by the product of cycles of prime lengths $2, 3, \ldots, p_i$ where $2 + 3 + \cdots + p_i \leqslant n < 2 + 3 + \cdots + p_{i+1}$. Unfortunately, however, no better bound is known for $G = S_n$ either.

We can do much better if the generators are chosen at random rather than adversarially.

Theorem 3.47. *Let σ, τ be two randomly selected permutations of a set of n elements and $G = \langle \sigma, \tau \rangle$.*

(a) *With probability* $1 - O(1/n)$, $A_n \leqslant G \leqslant S_n$ (Dixon 1969). (*The error term is from* Babai 1989.)

(b) *With probability* $1 - o(1)$, *the diameter of $\Gamma(G,S)$ is at most* $n^{(\ln n(1+o(1)))/2}$ (Babai and Hetyei 1992).

To appreciate the difficulty of determining the exact diameter with respect to a given set of generators, we mention two results on the computational complexity of this problem. For a permutation group $G \leqslant S_n$, it is NP-hard to determine the diameter of $\Gamma(G,S)$ even if G is an elementary abelian 2-group (Even and Goldreich 1981). For Cayley digraphs of permutation groups, it is a PSPACE-complete problem to determine the directed distance of a given pair of group elements (Jerrum 1985).

3.9. Automorphisms of maps

Note. In this section, both finite and infinite graphs will be considered. All surfaces (2-dimensional manifolds) considered are closed (without boundary) and compact, with the significant exception of the plane.

One has to make a distinction between the automorphism groups of ***graphs*** embeddable on a surface Σ and the automorphism groups of the ***maps*** defined by specific embeddings $X \subset \Sigma$.

Recall (cf. chapter 5) that a map is a graph X embedded on a surface Σ such that the components of $\Sigma \setminus X$, the *faces* of the map, are homeomorphic to an open disc. If the surface Σ is compact, X must be finite. *Map-automorphisms* preserve incidences between edges and faces in addition to those between edges and vertices. If v, e, and f denote the number of vertices, edges, and faces, resp., of a map on a compact surface then

$$v - e + f = \chi \tag{7}$$

where $\chi = \chi(\Sigma)$ denotes the Euler characteristic of Σ.

Recall that $\chi \leqslant 2$ is an integer. If Σ is orientable then χ is even; the quantity $g = 1 - \chi/2$ is the *genus* of Σ; and Σ is homeomorphic to the "sphere with g handles". If Σ is non-orientable then $g' = 2 - \chi$ is its non-orientable genus; and Σ is homeomorphic to the "sphere with g' crosscaps". Thus, orientability and the Euler characteristic characterize all compact surfaces up to homeomorphism.

The compact surfaces of non-negative Euler characteristic are the following: (a) orientable: the sphere ($\chi = 2$) and the torus ($\chi = 0$); (b) non-orientable: the projective plane ($\chi = 1$) and the Klein bottle ($\chi = 0$).

Map-automorphisms extend isomorphically to groups of homeomorphisms of Σ, and conversely: every finite group G acting on a compact surface Σ acts as a *vertex-transitive* group of automorphisms of some map. Unless Σ is the sphere, we may require in addition that every face has at least 3 sides. (For instance, if G is the trivial group and Σ is the torus, we shall have a single vertex with two loops, creating a single four-sided face.)

Each non-orientable surface Σ_1 has an orientable double cover Σ_2, of Euler characteristic $2\chi(\Sigma_1)$. The action of any group G on Σ_1 can be lifted isomorphically to an *orientable* action on Σ_2. The action of G on Σ_2 commutes with the sense-reversing "antipodal map" which switches the pairs of preimages of the covering map $\Sigma_2 \to \Sigma_1$, hence $G \times \mathbb{Z}_2$ acts on Σ_2. These facts follow from the elements of homotopy theory; cf. Tucker (1983, p. 96).

To understand finite group actions and maps on compact surfaces, we need to look at the three *natural geometries:* the sphere, the euclidean plane, and the hyperbolic plane. (These are the only simply connected 2-dimensional complete Riemannian manifolds of constant curvature.)

Let G be a finite group of homeomorphisms of the compact surface Σ of Euler characteristic χ. Then Σ admits a G-invariant Riemannian metric of constant curvature. The curvature will be positive, zero, or negative according to the sign of $\chi(\Sigma)$. This makes our surface Σ locally isometric to the corresponding natural geometry.

Moreover, if M is a vertex-transitive map on Σ, invariant under G, without one-sided or two-sided faces, then the metric can be chosen so as to make all egdes geodetic and all faces regular. (Cf. Jones and Singerman 1978, and the proof of Zieschang et al. 1980, Theorem 6.4.7.)

More about the groups and the maps on Σ can be found out by lifting them to $\tilde{\Sigma}$, the universal covering space of Σ, which is the natural geometry locally isometric to Σ.

We define a *crystallographic group* of a natural geometry as a discrete group of isometrics with compact fundamental domain[1].

Theorem 3.48. *Let G be a finite group acting on the compact surface Σ. Then G lifts to a crystallographic group $\tilde{G}$ of the natural geometry of its universal cover (sphere, euclidean plane, or hyperbolic plane).*

(Cf. Zieschang et al. 1980, Theorem 6.4.7.) The fundamental group $\pi_1(\Sigma)$ is normal in $\tilde{G}$ and $\tilde{G}/\pi_1(\Sigma) \cong G$.

An *Archimedean tiling* of a natural geometry is a map of which each face is a regular polygon and the map admits a vertex-transitive group of isometries.

Theorem 3.49. *A vertex-transitive map M on a compact surface Σ lifts to an Archimedean tiling of the natural geometry of the universal covering surface $\tilde{\Sigma}$.*

(Cf. the proof of Zieschang et al. 1980, Theorem 6.4.7.)

One can classify the crystallographic groups of the three natural geometries via canonical codes; each code is associated with a presentation in terms of generators and relations derived from a pair of dual maps. If two such groups are isomorphic as abstract groups then their isomorphisms are also geometrically realizable (Wilkie 1966, Macbeath 1967, cf. Zieschang et al. 1980, Theorems 4.5.6–4.7.1).

The crystallographic groups of the *sphere* are finite; they are listed in section 1.4. There are 18 individual types and two infinite one-parameter families of *vertex-transitive maps on the sphere*, corresponding to the Platonic and Archimedean solids and the families of prisms and antiprisms.

By the foregoing remarks, we obtain that the finite group actions on the *projective plane* are precisely the actions, on the pairs of antipodal points, of the finite rotation groups of the sphere.

Vertex-transitive maps on the projective plane correspond to centrally symmetric vertex-transitive maps on the sphere and are obtained from them by identifying antipodes.

When $\chi(\Sigma) = 0$, Theorem 3.48 relates G to the classical *crystallographic groups* of the euclidean plane. These were classified in the last century (Fedorov 1891). There are (up to natural equivalence) 17 of them (see Coxeter and Moser 1972, p. 44). Each crystallographic group G is equivalent to a group of isometries of the plane acting transitively on the points of a regular triangular, square, or hexagonal grid. It follows that the index of G is not greater than 12, 8, and 6, resp., in the full group of symmetries of the corresponding grid. Furthermore, G contains a normal subgroup N generated by two linearly independent translations, and the quotient G/N is a subgroup of the dihedral group of degree 6 or 4.

[1] This deviates from common usage in the hyperbolic case where compactness is usually not required.

Every *normal* subgroup H of G generated by two linearly independent translations gives rise to a unique action of G/H on the torus $\mathbb{R}^2/H$; this observation describes all finite group actions on the *torus*. (Note, in particular, that all these groups are solvable.)

The situation with the *Klein bottle* is similar except that the normal subgroup $H \lhd G$ must be generated by a translation and a *glide-reflection*, i.e. a translation followed by a reflection in an axis parallel to the direction of the translation. This implies severe restrictions on G. One can prove, in particular, that the square of any rotation belongs to H, and the subgroup $T \leqslant G$ of translations has a subgroup T_1 of index $\leqslant 2$ such that $T_1/(T_1 \cap H)$ is cyclic.

Corollary 3.50. (a) *Let G be a finite group acting on the torus. Then G has an abelian normal subgroup N with $\leqslant 2$ generators such that $G/N \cong \mathbb{Z}_k$ or D_k, $k = 6$ or $k \leqslant 4$.*

(b) *Let G be a finite group acting on the Klein bottle. Then $G \cong \mathbb{Z}_n$, D_n, $\mathbb{Z}_{2n} \times \mathbb{Z}_2$, or $D_{2n} \times \mathbb{Z}_2$.*

There are 11 types of Archimedean tilings of the euclidean plane (see fig. 1). Each of the tilings gives rise to a 2-parameter family of vertex-transitive *toroidal maps*.

The vertex-transitive maps without one-sided and two-sided faces on the Klein bottle form 13 families corresponding in different ways to 6 out of the 11 vertex-transitive euclidean tilings; each of them have "width 4" in the sense that all vertices belong to 4 straight lines parallel to the glide-reflection axis on the Klein bottle (Thomassen 1991, Babai 1991c). In a similar sense, the degenerate cases have "width" 2 or 1 and are also known.

When $\chi(\Sigma) < 0$, the finite groups acting on Σ are quotients of discrete subgroups of PGL(2, $\mathbb{R}$), the isometry group of the hyperbolic plane. A classical theorem of Hurwitz (1893) indicates a drastic change compared to the case $\chi \geqslant 0$.

Theorem 3.51 (Hurwitz). *If the finite group G acts on the compact surface Σ of Euler characteristic $\chi < 0$ then $|G| \leqslant 84|\chi|$.*

For a proof when Σ is orientable, see, e.g., Gross and Tucker (1987 p. 496). The general case follows by the foregoing remarks. There are infinitely many values of χ where the bound $84|\chi|$ is attained (Conder 1980).

The following is a combinatorial generalization of Hurwitz's Theorem. With each vertex of a map we associate a cyclically ordered list containing the number of sides of each face incident with the vertex. We call a map *semiregular* if the cyclic list associated with each vertex is the same (up to inversion).

Theorem 3.52 (Babai 1991c). *Let M be a semiregular map on a compact surface of Euler characteristic $\chi < 0$. Then M has at most $84|\chi|$ vertices.*

Each homeomorphism ψ of a compact orientable surface Σ of genus g induces an automorphism ψ_* of the first homology group $H_1(\Sigma) \cong \mathbb{Z}^{2g}$ which preserves a skew-symmetric bilinear form $H_1(\Sigma) \times H_1(\Sigma) \to \mathbb{Z}$, defined by the *intersection numbers*

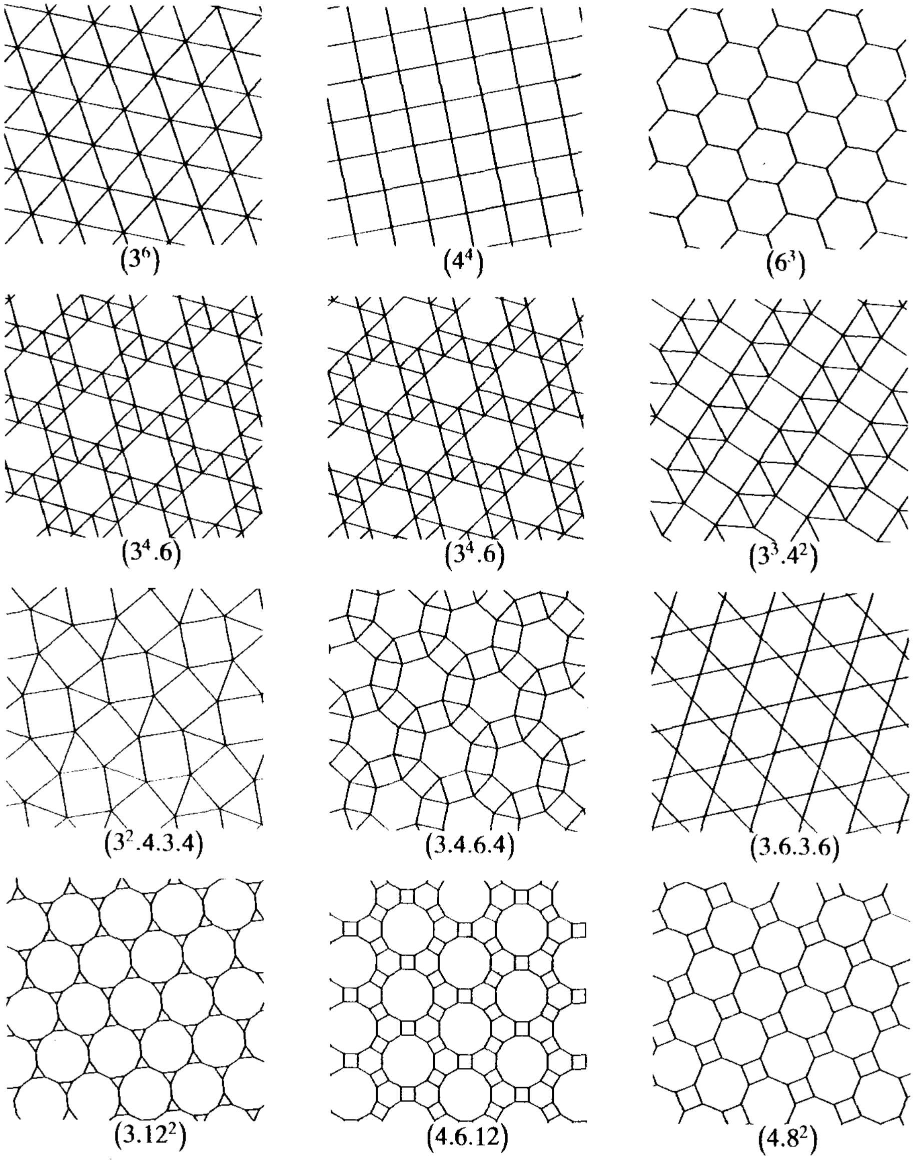


Figure 1. The eleven types of Archimedean tilings of the euclidean plane (one of them shown in two mirror-symmetrical forms). After Grünbaum and Shephard (1981 p. 144).

of curves, cf. Zieschang et al. (1980, Proposition 3.6.3). Another result of Hurwitz (1893) states that if ψ has finite order and $g \geqslant 2$ then $\psi_* \neq 1$. Consequently, if G is a finite group of homeomorphisms of Σ then G is isomorphic to a subgroup of $\mathrm{Sp}(g, \mathbb{Z})$, the group of $2g \times 2g$ integral symplectic matrices (cf. Zieschang et al. 1980, Corollary 4.15.3, Biggs 1972). This result also holds for the homology groups mod n for any $n \geqslant 3$ (Serre in 1960, cf. Zieschang et al. 1980, Corollary 4.15.15).

3.10. Embeddings on surfaces, minors

Note. All graphs and groups in this section are finite, except in the last paragraphs (beginning after Theorem 3.59).

Now we turn to the question of classifying the connected vertex transitive graphs embeddable on a given surface. If an embedding of the graph X on the surface Σ creates a map and all automorphisms of X extend to map-automorphisms then we call the embedding *automorphic*. The main difficulty is that embeddings are seldom automorphic. Some of the most surprising results in the area infer the existence of automorphic embeddings from seemingly unrelated asymptotic combinatorial assumptions.

Apart from cycles, vertex p-transitive graphs have degree $\geqslant 3$ and are therefore 3-connected. For planar graphs this implies unique embeddability on the sphere, hence those embeddings are automorphic, and the list of 18 types plus two infinite one-parameter families mentioned above applies.

For no other surface Σ have the vertex-transitive graphs embeddable on Σ been fully classified. Interest in embedding Cayley graphs on surfaces has been motivated since the last century by the following observation: Cayley graph embeddings help in finding *presentations* (generators and relations) for G.

Proposition 3.53. *Let $X = \Gamma(G, S)$ be embedded on Σ. Let the cycles $C_1, \ldots, C_m$ of X through 1 generate the fundamental group of Σ. Let further $D_1, \ldots, D_{f-1}$ denote all but one of the fundamental cycles (face boundaries) of the embedding. Then the C_i and the D_j, regarded as words in the symbols $S \cup S^{-1}$, form a complete set of relations defining G. If the map is vertex-transitive, only those D_j passing through 1 have to be taken.*

Proof. Every cycle in the Cayley graph indicates a valid relation among the generators. We have to show that every cycle A represents a consequence of the relations listed. We may assume A passes through 1. Then A, as a path on Σ, is homotopic to some product P of the C_i. It follows that AP^{-1} is contractible and therefore representable as a product of the D_j. □

Maschke (1896) determined all planar minimal Cayley diagrams. (Recall that we call $\Gamma(G, S)$ and $\Gamma_c(G, S)$ *minimal* if S generates G with no redundant elements.) Nonplanar toroidal minimal Cayley diagrams have been classified by Proulx. Her list contains 11 infinite classes with two generators, 9 infinite classes with 3 generators, 1 infinite class with 4 generators, and 9 sporadic cases (8 with 2 generators, 1 with three). We state the main consequence.

Theorem 3.54 (Proulx 1977). *All but 3 of the groups admitting a toroidal but no planar Cayley graphs are quotient groups of euclidean 2-dimensional crystallographic groups and therefore they actually admit automorphically embedded toroidal Cayley graphs.*

The precise set of 3 exceptions (of orders 24, 48 and 48) has been determined by Tucker (1984). Using in great detail Proulx's analysis, Tucker went on to proving an extension of Hurwitz's theorem to Cayley graphs embeddable on surfaces of negative Euler characteristic.

Theorem 3.55 (Tucker 1984). *Let G be a group of order n and $\Gamma(G,S)$ a minimal Cayley graph of G, embeddable on a surface Σ of Euler characteristic $\chi < 0$ but not embeddable on the torus. Then $|G| \leqslant 84|\chi|$.*

We indicate some of the basic tricks of the Proulx–Tucker theory on a very simple special case.

Proposition 3.56. *Let G be a group of order n where $\mathrm{GCD}(n,6) = 1$. Assume G has a minimal Cayley graph X embeddable on a surface Σ of Euler characteristic χ. If $n > -5\chi$, then G is abelian with two generators and X is toroidal.*

Proof. Let $X = \Gamma(G,S)$ have degree $d \geqslant 3$. Since G has no elements of order 3, the girth of G is $\geqslant 4$. Now X has n vertices, $e = nd/2$ edges, and $f \leqslant nd/4$ faces. Substituting into the Euler equation (7) we obtain

$$n(1 - d/4) \geqslant \chi.$$

If $d \geqslant 5$, we infer $n \leqslant -4\chi$. If $d = 3$ then one of the generators would have to be an involution, impossible. The only remaining case is $d = 4$; hence $e = 2n$ and S consists of 2 elements: $S = \{a,b\}$.

Assume first that the girth of X is $\geqslant 5$. Let f_i denote the number of i-sided faces. We then have $2e = \sum_{i \geqslant 5} i f_i$ and $f = \sum_{i \geqslant 5} f_i \leqslant 2e/5 = 4n/5$. Hence

$$\chi = n - e + f \leqslant n - 2n + 4n/5 = -n/5.$$

We conclude that $n \leqslant -5\chi$, thus finishing this case.

We may henceforth assume that the girth of X is 4. By minimality, the implied relation of length 4 must be of one of the following types: (a) $a^4 = 1$; (b) $abab = 1$; (c) $aba^{-1}b = 1$; (d) $a^2b^2 = 1$; (e) $aba^{-1}b^{-1} = 1$.

Since G has odd order and S is minimal, only case (e) can actually occur. But then, G is abelian with two generators, hence it is toroidal. □

While the arguments that count degrees and use the Euler equation generalize to arbitrary vertex transitive graphs, the "relation chasing" that concluded the proof has no analogue. Arguments of a more geometric flavor, however, yield the following.

Theorem 3.57 (Thomassen 1991, Babai 1991c). *There exists a function f such that every connected vertex-transitive graph X with more than $f(\chi)$ vertices and embeddable on a surface of Euler characteristic χ admits an embedding as a vertex-transitive* map *on a surface of nonnegative Euler characteristic.*

The function f is bounded by $c|\chi|$ where c is an absolute constant (Thomassen 1991). With the exception of 4 families of "crossed stripe-like" graphs (Babai 1991c), the embeddings guaranteed by the theorem are *automorphic.*

Embeddings of specific Cayley graphs, in particular of complete graphs viewed as Cayley graphs of cyclic groups, have been studied extensively. The original motivation for this was the solution, due mainly to Ringel and Youngs (1968), of the Heawood map color conjecture. (For details and references, we refer to the monograph of Gross and Tucker 1987.) Subsequently, the following concept gained popularity.

Definition. The *genus* of a (finite) group G is the minimum of the genera of those orientable compact surfaces Σ on which some connected Cayley graph of G is embeddable. The *non-orientable genus* of G is the minimum of $(2 - \chi(\Sigma))$ over the corresponding not necessarily orientable surfaces Σ.

Both the orientable and non-orientable genera are monotone for subgroups (Babai 1977a). (This follows immediately from Proposition 3.58 below.) It is an open question whether or not the same holds for quotient groups, as conjectured by White (1973). Jungerman and White (1980) were able to determine the precise genus for surprisingly large classes of abelian groups, demonstrating that those groups admit embeddings with quadrilateral faces. The situation becomes more complicated when $\mathbb{Z}_3$ factors are present and triangular faces may arise.

Contractions tend to simplify the topological characteristics of a graph. Significantly, they can be related to group actions.

Proposition 3.58 (The "Contraction lemma"). *If the group G acts semiregularly on the connected graph X then X has a contraction to some Cayley graph of G. In particular, if $G \leqslant H$ then every Cay*
Cayley graph of G (Serre 1977, Babai

(Semiregular action means that the s
immediate consequence is the Nielsen
groups are free.

The *Hadwiger number* of a (not ne
of those values k such that some con
The Hadwiger number of graphs embe
for the torus, this bound is 7). The co
give an asymptotic classification of all
Hadwiger numbers.

Theorem 3.59 (Babai 1994). *There exi*
nected vertex-transitive graph X of H

cf. G
a graph ngly
The s (iv) but
orbits. First
automorphism
constructions rela
where $n = |G|$ and
is a consequence of th V. The set of graphs
described; their number

admitting an automorphic embedding on the torus, or (b) *ring-like in the following sense: $V(X)$ has a partition $(V_0, \ldots, V_{m-1})$ into blocks of imprimitivity such that* (b1) $|V_i| < f(k)$; (b2) *if there is an edge between V_i and V_j then $|i - j| < f(k)$ or $m - |i - j| < f(k)$*; (b3) *the action of* $\operatorname{Aut} X$ *on the set of blocks is either cyclic or dihedral.*

The proof requires the study of the *local structure* of the graphs via an infinite vertex-transitive "limit graph" (cf. Babai 1991c) and distinguishes cases according to the number of *ends* of the limit graph, using Proposition 3.31. The case of infinitely many ends is disposed of using a sphere packing argument (Babai 1991c) motivated by Thomassen's proof that graphs of degree $\geqslant 3$ and large girth have large Hadwiger number (Thomassen 1983)

The case of two ends yields ring-like graphs, using Dunwoody's theorem 3.37. The hard case is when the limit has a single end. The analysis requires the following result.

Theorem 3.60 (Thomassen 1992). *Let X be an infinite locally finite connected vertex-transitive graph with a single end. If X has finite Hadwiger number then X is planar.*

Such an infinite graph, then, can be shown to have a natural associated geometry (along the lines of Theorem. 3.49, cf. Babai 1994):

Theorem 3.61. *Let X be an infinite locally finite connected vertex-transitive planar graph with a single end. Then X has an automorphic embedding as a tiling of the euclidean or hyperbolic plane.*

Returning to the sketch of the proof of Theorem 3.59, we observe that euclidean tilings give rise to toroidal graphs. Hyperbolic tilings lead to finite graphs of large Hadwiger number, via another sphere packing argument, using the elements of hyperbolic geometry.

3.11. *Combinatorial group theory*

Combinatorial group theory investigates presentations of groups defined in terms of generators and relations. Typical constructions in this field are the free product with amalgams, and HNN-extensions (cf. section 3.7). One of the classical results is the Nielsen–Schreier theorem that *every subgroup of a free group is free.* This, incidentally, follows immediately from the Contraction lemma (Proposition 3.58). Indeed, among the groups with no elements of order 2, precisely the free groups have trees for Cayley graphs; and a contraction of a tree is a tree again.

There is no way we could do justice to this vast area here; the reader is referred to the monographs by Coxeter and Moser (1972), Magnus et al. (1966), Lyndon and Schupp (1977), Serre (1980), Dicks and Dunwoody (1989). The elementary graph theoretic approach to classical subgroup theorems is emphasized in Imrich's (1977) friendly notes. When proving subgroup theorems such as Kurosh's theorem stated at the end of this section, the basic geometric object to consider is the quotient of a Cayley diagram of the group G by the action of the subgroup H

(Schreier coset diagram). An example of an interesting result in this area proved by an elementary graph theoretic argument is *Howson's theorem:* the intersection of two finitely generated subgroups of a free group is finitely generated (Imrich 1977, Dicks and Dunwoody 1989, I.8, Tardos 1992).

Some of the results mentioned earlier in this chapter belong to combinatorial group theory (e.g., Proposition 3.53 or Dunwoody's theorem 3.36).

A relatively recent highlight is the *Bass–Serre theory* of group actions on trees. They introduce a construction called a *graph of groups* in which a group $G(v)$ is assigned to each vertex v of a directed graph and a subgroup $G(v,w) \leqslant G(v)$ to every directed edge (v,w), along with an injective homomorphism $t_{v,w}: G(v,w) \to G(w)$. The *fundamental group* of a graph of groups is defined as a group generated by the disjoint union of the $G(v)$ along with one symbol $t_{v,w}$ for every edge (v,w), subject to the relations defining $G(v)$ for each v, and the relations $t_{v,w}^{-1} g t_{v,w} = g^{t_{v,w}}$ for each edge (v,w) and element $g \in G(v,w)$. (Note that therefore $g \in G(v)$ and $g^{t_{v,w}} \in G(w)$.) Moreover, we select an arbitrary maximal subtree of the graph, and set $t_{v,w} = 1$ for every edge (v,w) in the tree.

Observe that if the graph consists of a single directed edge (v,w) then the fundamental group will be the free product of $G(v)$ and $G(w)$, with the subgroup $G(v,w)$ amalgamated. If the graph has a single vertex v with a loop (v,v) then the fundamental group is the HNN-extension $(G(v), G(v,v), t_{v,v})$. This is a restatement of Theorems 3.34 and 3.35.

Let now G be a group acting on a tree T without inverting edges. Then the *Bass–Serre structure theorem* asserts that G is isomorphic to the fundamental group of a graph of groups, where the graph is the quotient graph of T by the action of G (Serre 1980, section I.5.4, section I.4).

Among the immediate consequences is Kurosh's classical subgroup theorem, asserting that a subgroup of a free product of the groups G_i is a free product of a free group and conjugates of subgroups of the G_i.

3.12. *Eigenvalues*

Let $\alpha: G \to \mathbb{C}$ be a function, and consider the $n \times n$ matrix $A = (a_{g,h})$, whose rows as well as columns are labeled by the elements of G (in the same order, $n = |G|$), and

$$a_{g,h} = \alpha(gh^{-1}).$$

We can think of α as a "color assignment" to the elements of G; thus A is the adjacency matrix of a Cayley color diagram. We call A a G-circulant, since in the case $G = \mathbb{Z}_n$ we obtain precisely the circulant matrices.

In the circulant case, det A has a well-known expansion into linear factors. Let ω denote a primitive nth root of unity; then the vectors $w_i = (1, \omega^i, \omega^{2i}, \ldots, \omega^{(n-1)i})$ form a system of orthogonal eigenvectors of A, with corresponding eigenvalues

$$\lambda_i = \sum_k \alpha(k)\omega^{ik}. \tag{8}$$

The determinant of A is $\prod_i \lambda_i$.

Examining the expansion of det A for the dihedral groups, Dedekind noticed that (viewing each value $\alpha(g)$ as an independent variable) most irreducible factors were no longer linear but quadratic, and called on Frobenius in a letter to investigate the general case. Frobenius soon found a wealth of structure; his paper "Über die Primfactoren der Gruppendeterminante", presented to the Prussian Academy of Sciences in 1896, laid the foundations of character theory for nonabelian groups.

A consequence of this theory is, that, denoting the dimensions of the irreducible characters of G by $n_1, \ldots, n_h$ (h is the number of conjugacy classes in G; and $\sum n_i^2 = n$), the eigenvalues of any G-circulant can be assigned to irreducible characters in the following way: n_i^2 eigenvalues correspond to character χ_i; these fall into n_i equal groups, and all the n_i eigenvalues within a group are equal. Moreover, the sum of the potentially different n_i eigenvalues (one from each group of n_i) belonging to χ_i is

$$\lambda_{i,1} + \cdots + \lambda_{i,n_i} = \sum_{g \in G} \alpha(g)\chi_i(g). \tag{9}$$

(See Babai 1979d.) In particular, if G is abelian, then each $n_i = 1$, and the expression simplifies to

$$\lambda_i = \sum_{g \in G} \alpha(g)\chi_i(g), \tag{10}$$

a direct generalization of the circulant case (eq. (8)).

As an example, let $X = X(n,k)$ denote the distance-k graph of the n-dimensional cube. Let A be an n-set and let us represent the elements of the n-cube by subsets of A. With the operation of symmetric difference, this set is the elementary abelian group $\mathbb{Z}_2^n$ and $X = \Gamma(\mathbb{Z}_2^n, S_k)$ where S_k is the set of all k-subsets of A. Characters $\chi_T : \mathbb{Z}_2^n \to \{\pm 1\}$ are associated with subsets $T \subseteq A$ via the rule $\chi_T(B) = (-1)^{|T \cap B|}$. The corresponding eigenvalue of X is $\lambda_T = \sum_{|B|=k} (-1)^{|T \cap B|} = K_k(|T|)$, where

$$K_k(x) = \sum_{i=0}^{k} (-1)^i \binom{x}{i}\binom{n-x}{k-i} \tag{11}$$

is the Krawtchouk polynomial (cf. Bannai and Ito 1984, section 3.2).

A more general class of G-circulants, admitting an explicit expression of their eigenvalues, are obtained when α is a *class function*, i.e. α is constant on conjugacy classes. (In other words, $\alpha(gh) = \alpha(hg)$ for every $g, h \in G$.) In this case all the n_i^2 eigenvalues belonging to χ_i are equal, and hence their common value is $\lambda_i = (1/n_i)\sum_{g \in G} \alpha(g)\chi_i(g)$ and the matrix A is diagonalizable (via a unitary transformation).

It follows from the above that if the set S of generators is closed under conjugation then the adjacency matrix of the Cayley digraph $\Gamma(G, S)$ is diagonalizable. This is not true for general S; Godsil (1982) has shown that the minimal polynomial of any integral matrix divides the minimal polynomial of some Cayley digraph.

Cayley graphs of cyclic groups of prime order are determined up to isomorphism by their characteristic polynomials (Elspas and Turner 1970). This is not true for general groups; families of isospectral Cayley graphs of the dihedral groups of all odd prime degrees are exhibited in Babai (1979d).

The results discussed above belong to the *harmonic analysis* over G. For an exposition and a variety of applications (especially to random walks), see Diaconis (1988, chapter 3; 1989), Chillag (1988).

For the extensive literature on the harmonic analysis over locally finite infinite graphs we refer to the survey Mohar and Woess (1989).

4. The representation problem

The material of this section is covered in greater detail in the survey paper by Babai (1981b) where additional references and in many cases complete proofs can be found.

4.1. Abstract representation; prescribed properties

In this section we consider the following type of problem: given a group G find a graph X (or a block design, a lattice, a ring, etc.) such that the automorphism group $\mathrm{Aut}(X)$ is *isomorphic* to G. Such an object X will be said to *represent* the group G. A class $\mathscr{C}$ of objects is said to represent a class $\mathscr{G}$ of groups if, given $G \in \mathscr{G}$ there exists $X \in \mathscr{C}$ such that $\mathrm{Aut}(X) \cong G$. We call $\mathscr{C}$ *universal*, if every group is represented by $\mathscr{C}$. We say that $\mathscr{C}$ is *finitely universal* if every finite group occurs among the groups represented by *finite* members of $\mathscr{C}$.

The natural question, which groups are represented by graphs, was stated by König (1936), and soon answered by Frucht.

Theorem 4.1 (Frucht 1938). *Given a finite group G there exists a finite graph X such that* $\mathrm{Aut}(X) \cong G$. *In other words, graphs are finitely universal.*

Frucht's proof has been reproduced in several texts (Ore 1962, Harary 1969, Lovász 1979a, Bollobás 1979). The idea is (i) to observe that the automorphism group of the (colored, directed) Cayley diagram of G with respect to any set of generators is isomorphic to G; (ii) to get rid of colors and orientation by replacing colored arrows by appropriate small asymmetric (automorphism free) gadgets.

The next problem was to find subclasses of graphs and classes of other (combinatorial, algebraic, topological) objects that are universal. This direction was initiated by Frucht and Birkhoff. Frucht (1949) proved that *trivalent graphs are finitely universal*. It is immediate from Theorem 4.1 that posets are finitely universal. Since posets are strongly reconstructible from their lattice of ideals (as the poset of join-irreducible elements), it follows that *distributive lattices are finitely universal* (as well as universal, Birkhoff 1945).

These results already foreshadow the lopsidedness of later developments. Take almost any "reasonably broad" class of combinatorial or algebraic structures; the class will be universal. (Groups, planar graphs are notable exceptions.) This "universality phenomenon" was first indicated by Sabidussi (1957); he proved that Hamiltonicity, k-regularity, k-connectedness are all compatible with any prescribed automorphism group. Universality results in topology and algebra were inspired by De Groot's (1958, 1959) papers, where topological spaces and commutative rings were shown to be universal. A surprisingly strong version of the latter result was given by E. Fried and J. Kollár.

Theorem 4.2 (Fried and Kollár 1978, 1981). *Every group is the automorphism group of a field. Every finite group is the automorphism group of an algebraic number field.*

(Algebraic number fields are finite extensions of $\mathbb{Q}$.) The proof takes a graph X with the given automorphism group and encodes it into a field (not without ingenuity). This is the basic scheme of most universality proofs.

The extensions constructed by Fried and Kollár are not normal. Therefore their result does not bear on the inverse problem of Galois theory (represent a given group as a Galois group over a given field; notably, over $\mathbb{Q}$). We note in passing that the inverse problem has had its renaissance in the past decade, inspired by Thompson's (1984) new approach (cf. Feit 1989, Matzat 1987, several articles in Aschbacher et al. 1985). One of Thompson's corollaries states that the Monster, the largest sporadic simple group, is a Galois group over $\mathbb{Q}$.

Of the numerous combinatorial universality results, let me quote two of the more surprising ones.

Theorem 4.3 (Mendelsohn 1978a,b). *Every finite group is the automorphism group of* (a) *a finite Steiner triple system and a finite Steiner quadruple system;* (b) *a finite strongly regular graph.*

Universality proofs usually require *reconstruction arguments*. To illustrate this point, we deduce Mendelsohn's result (b) from (a). Let X be a Steiner triple system with the prescribed automorphism group G. Take its line graph $L(X)$. $L(X)$ is strongly regular, and, according to Theorem 1.12, X is strongly reconstructible from $L(X)$, assuming X has > 15 vertices. In particular, $\mathrm{Aut}(X) \cong \mathrm{Aut}(L(X))$ (Corollary 1.13(a)). □

The automorphism group is very sensitive to slight changes in the graph. It is known, for instance, that for any pair of groups G and H there exists a graph X and an edge $e \in E(X)$ such that $\mathrm{Aut}(X) \cong G$ and $\mathrm{Aut}(X \setminus e) \cong H$ (Babai, see Lovász 1979a, Example 12.11).

It is typical for universality proofs that the group structure plays a small role. The extent to which group structure can be ignored is demonstrated by generalizations to prescribability of semigroups of *endomorphisms* and even categories, pioneered by the Prague category theory school, especially Pultr and Hedrlín. A *homomorphism* of the graph X to the graph Y is an adjacency preserving map $V(X) \to V(Y)$. Note that non-adjacent vertices may have the same image. *Endomorphisms* of a graph X are homomorphisms $X \to X$. They form the monoid

End(X). (Monoid = semigroup with identity.) The basic result is that *every monoid is the endomorphism monoid of some graph* (finite graphs for finite monoids) (Hedrlín, Pultr, Vopeňka). By encoding graphs, many classes of algebraic and topological structures have been shown to have the same property (see the monograph by Pultr and Trnková 1980). A nice introduction to the subject is Hedrlín and Lambek (1969).

Universality-type results are known for some classes of structures that are clearly not universal.

Theorem 4.4. (a) *The automorphism groups of (finite) tournaments have odd order; and every finite group of odd order is represented by a tournament* (Moon 1964).

(b) *G is the automorphism group of a switching class of tournaments if and only if its Sylow 2-subgroups are cyclic or dihedral* (Babai and Cameron 1994a). (*Two tournaments T_1, T_2 on the common vertex set V are switching equivalent if V can be partitioned into two classes such that one obtains T_2 from T_1 by reversing all edges between the two classes. This equivalence relation divides the set of tournaments on V into switching classes.*)

(c) *Denote by Γ_d the class of groups G with a subgroup chain $G = G_0 \geqslant G_1 \geqslant \cdots \geqslant G_m = 1$ such that $|G_{i-1} : G_i| \leqslant d$ for every i. If X is a connected regular graph of degree $d+1$ then the stabilizer of an edge in X belongs to Γ_d; and every group in Γ_d can be represented this way* (Babai and Lovász 1973).

It is an open problem to show that the *finite projective planes are not finitely universal*, i.e., not every group is isomorphic to the automorphism group of a finite projective plane. Indeed it seems plausible that most finite groups cannot act on a finite projective plane (as a subgroup of the automorphism group), but no group has been ruled out so far. Hering (1967) proved that for $n \equiv 3 \pmod 4$, any 2-group acting on a projective plane of order n must be cyclic, a (generalized) quaternion group, a dihedral group, or a quasidihedral group.

4.2. Topological properties

Topological properties of a graph (embeddability on a surface, excluded minors) do restrict the abstract group of automorphisms and thus offer a welcome source of connections between the structure of groups and the graphs representing them. The following general non-universality result says that prescribed automorphism groups force arbitrary minors to occur.

Theorem 4.5 (Babai 1974a). *Given a finite graph Y there exists a finite group G such that every graph X with* Aut(X) $\cong$ *G has Y as a minor.*

Let $\mathscr{C}(Y)$ be the class of finite graphs without Y as a minor. It is expected that the finite groups represented by $\mathscr{C}(Y)$ have a very restricted structure. In particular, it is conjectured that the list of nonabelian finite simple groups represented by $\mathscr{C}(Y)$ is finite (cf. Babai 1981b).

Excluding a *topological subgraph* is, in general, less restrictive than excluding a minor; indeed even the exclusion of vertices of degree $\geqslant 4$ does not restrict the abstract automorphism group. However, prescribed endomorphism monoids do force arbitrary topological subgraphs: another strong non-universality result.

Theorem 4.6 (Babai and Pultr 1980). *Given a finite graph Y there exists a finite monoid M such that for every graph X, if* $\mathrm{End}(X) \cong M$ *then X contains a subdivision of Y.*

4.3. Small graphs with given group

The number of orbits is a measure of symmetry. It is natural to ask, how symmetrical the graphs representing a given group can be. By the *orbits of a graph*, we mean the orbits of its automorphism group on the vertex set. Edge-orbits are orbits on the edge set.

With three exceptions, every finite group can be represented by a graph with $\leqslant 2$ orbits (Babai 1974b). (The exceptions are the cyclic groups of orders 3, 4, and 5.)

Most groups even admit a *representation by a vertex-transitive graph.* Nowitz (1968) and Watkins (1971) described an infinite family of groups without a vertex-transitive representation (abelian groups of exponent greater than 2, and generalized dicyclic groups). Hetzel (1976) and Godsil (1981a) proved that apart from these, there is only a finite number of additional exceptions, each of order $\leqslant 32$. Godsil (1979) extended this result to finitely generated infinite groups.

A *graphical regular representation* (GRR) of a group G is a graph X such that $\mathrm{Aut}(G)$ is regular and isomorphic to G. In other words, X is a Cayley graph of G without "extra" automorphisms (all automorphisms correspond to right translations, cf. section 3.1).

The graphs Hetzel and Godsil construct are actually GRRs and the result stated constitutes the full solution of the GRR problem: the characterization of all finite groups which admit a GRR. For certain classes of groups G, including all nonabelian nilpotent groups of odd order, one can actually show that *almost all Cayley graphs of G are GRRs* (Babai and Godsil 1982). (To obtain a random Cayley graph $\Gamma(G,S)$, one chooses a symmetrical set $S = S^{-1} \subseteq G$ at random.)

The analogous problem for digraphs is easier: with 5 exceptions, all groups (finite or infinite) have a *digraphical regular representation* (Babai 1978c, 1980a). (The exceptions are the elementary abelian groups of orders 4, 8, 9, 16, and the quaternion group of order 8. For infinite groups, the proof employs infinite Ramsey theory, cf. chapter 42.) A consequence is that every infinite group can be represented by a graph X with 3 orbits.

The situation is quite different when we wish to minimize the number of *edge-orbits*. First of all, if X is a graph representing the group G with a *semiregular* automorphism group (as has been the case so far in this section as well as in most constructions related to section 4.1) then the number of edges of X is at least $nd/2$, where $n = |G|$ and d is the minimum size of a symmetrical set of generators. (This is a consequence of the Contraction lemma 3.58.)

Let $e(G)$ denote the *minimum number of edges* and $m_e(G)$ the *minimum number of edge orbits* of the graphs representing G. Clearly $e(G)/|G| \leqslant m_e(G)$ and it is easy to see that $m_e(G) < C \log|G|$. Two natural questions arise: (a) is $m_e(G)$ bounded? (b) Is $e(G)/|G|$ bounded? We now have a fairly complete answer to both questions.

Theorem 4.7. (a) (Babai and Goodman 1991) *For all finite groups, $e(G)/|G| < 500$.*

(b) (Babai, Goodman and Lovász 1991) *If a finite group is generated by k abelian subgroups then $m_e(G) \leqslant Ck$ for some absolute constant C. (Note that, e.g., any direct product of finite simple groups is generated by $k = 2$ abelian subgroups.)*

(c) (Goodman 1993) *There is a constant $c > 0$ such that for infinitely many finite groups G, $m_e(G) > c\sqrt{\log|G|}$.*

(d)(S. Thomas 1987) *Assuming the generalized continuum hypothesis, for every successor cardinal κ there exists a group G of order κ such that $m_e(G) = \kappa$.*

The proof of (b) is related to a generalization of a result by Gel'fand and Ponomarev (1970) that the *subspace lattice* of a vector space of finite dimension $\geqslant 3$ over a prime field *is generated by* 4 *subspaces*. The proof of (c) has a curious nonconstructive element: certain p-groups of class two, demonstrating the lower bound, are shown to exist by a probabilistic (counting) argument. No explicit family of finite groups with unbounded $m_e(G)$ is known. The groups required for the proof of (d) are *Jónsson groups*, i.e. groups having no proper subgroups of their own cardinality. Shelah (1980) proved the existence of such groups for successor cardinals under GCH. Without GCH, no proof is known of the conjecture that $m_e(G)$ can be an arbitrarily large cardinal. □

Some classes of groups are represented by drastically smaller graphs. This is clear for the symmetric groups (graphs of order k represent the group of order $n = k!$), but less evident for the alternating groups (graphs of order $< 2^{k+1}$ represent the alternating group of order $k!/2$). Liebeck (1983) determines the *exact* minimum order of graphs representing the alternating group A_k for sufficiently large k (e.g., for $k \equiv 0, 1 \pmod 4$ he finds this minimum to be $2^k - k - 2$). (For small k, there are surprises, e.g., $A_8 \cong \mathrm{PSL}(4,2)$ is the automorphism group of a 30-vertex graph: the incidence graph of the projective geometry PG(3, 2).) Liebeck also gives strong lower bounds for the minimum order of graphs representing 3 types of classical simple groups (linear, orthogonal, unitary).

We mention related *open problems*. Let G be a group of order n. It follows from part (a) of the above theorem that G can be represented by a lattice of size $O(n)$. Can G be represented (i) by a *lattice with a bounded number of orbits?* Can G be represented by a *polynomial size* ($n^{O(1)}$) (ii) Steiner triple system, (iii) strongly regular graph, (iv) modular lattice? We conjecture the negative answer to (iv) but positive answers to (i)–(iii).

4.4. The concrete representation problem, 2-closure

Let $G \leqslant \mathrm{Sym}(V)$ be a permutation group, acting on the set V. The set of graphs $X = (V, E)$ admitting G as a subgroup of $\mathrm{Aut}(X)$ is easily described; their number is 2^k where k is the number of orbits of the induced action of G on the set of $\binom{|V|}{2}$ pairs.

The *concrete representation problem* asks if $G = \operatorname{Aut} X$ for some graph (digraph, etc.) with vertex set V. This problem is very difficult in general, as the case of regular permutation groups (the GRR problem, section 4.3) has demonstrated. But there is a simple necessary condition.

Let us consider the colored complete directed graph W with vertex set V obtained from G as follows. Two pairs of vertices receive the same color if and only if they belong to the same orbit of the induced action of G on $V \times V$. Vertex v receives the color of the pair (v, v). This is the *coherent configuration* corresponding to the group G. We define W^* to be the undirected version of W: unordered pairs receive colors.

We call $\operatorname{Aut}(W)$ the *2-closure* of G, and $\operatorname{Aut}(W^*)$ the *2^*-closure*. In other words, the 2-closure of G is the largest subgroup of $\operatorname{Sym}(V)$ with the same orbits on $V \times V$; and the *2^*-closure* the largest subgroup with the same orbits on points and on unordered pairs.

The group G is *2-closed* if G is equal to its 2-closure; 2^*-closed groups are defined analogously. A group is 2-closed if and only if it is the automorphism group of a colored directed graph; and 2^*-closedness corresponds to colored undirected graphs. These are thus necessary (but not sufficient) conditions for the group to be the automorphism group of a digraph (graph).

All regular permutation groups are 2-closed. Not all of them are 2^*-closed; the exceptions are precisely the abelian groups of exponent greater than two and the generalized dicyclic groups (Nowitz 1968, Watkins 1971, Babai 1977b).

For transitive permutation groups G, Godsil (1981b) gives further necessary conditions which for some class of nilpotent groups turn out also to be sufficient.

It is an interesting question, *how far the 2-closure* $\mathrm{cl}_2(G)$ *is from a group* G. Liebeck, Praeger and Saxl (1988) investigate this for the case when G is primitive and almost simple, i.e. $L \lhd G \leqslant \operatorname{Aut}(L)$ for some simple group L. If G is 2-transitive then $\mathrm{cl}_2(G) = \operatorname{Sym}(V)$; but the gap is much smaller in all other cases. Indeed Liebeck et al. (1988) find that $\mathrm{cl}_2(G)$ *normalizes* G, with the exception of six sporadic cases (the largest degree occurring in a representation of degree 276 of the Mathieu group M_{24}) plus two surprising infinite families of unbounded ranks with socles $L = G_2(q)$ and $\Omega_7(q)$, resp.

The notion of 2-closure as a tool in the study of permutation groups was introduced by I. Schur, see Wielandt (1969).

A maximal, not doubly transitive subgroup of S_n is necessarily 2-closed. This observation was used by L. A. Kaluzhnin and M. H. Klin in 1972 (cf. Klin et al. 1991) to give elementary proofs of the maximality of several classes of primitive groups, including the induced action of S_m on k-tuples ($n = \binom{m}{k}$), with some restrictions on (m, k). (For a complete study of this question via the classification of finite simple groups, see Liebeck, Praeger and Saxl 1987a.)

It is natural to ask which permutation groups arise as the automorphism groups of a hypergraph. If the sizes of the edges are not restricted, we have a nearly complete answer for primitive groups. Obviously, $A_n \neq \mathrm{Aut}(X)$ for any hypergraph X on n vertices. Apart from the alternating groups and an (unknown) finite family of other exceptions, *all primitive groups G occur as* Aut(X) *for some edge-transitive hypergraph* (Babai and Cameron 1994b). Exceptions include all set-transitive groups: the Frobenius group of order 20 ($n = 5$), PGL(2,5) ($n = 6$), PGL(2,8), PΓL(2,8) ($n = 9$). Another exception is the Frobenius group of order 21 ($n = 7$).

5. High symmetry

As in the Introduction, we shall use the abbreviation CFSG to indicate the classification of finite simple groups. CFSG has played a decisive role in the recent development of some of the subjects to be discussed below; we shall try to indicate where this is the case.

5.1. Locally s-arc-transitive graphs

All graphs in this section will be assumed finite and *connected.*

Let $s \geqslant 1$. An *s-arc* starting at a vertex v_0 in a graph X is a sequence $(v_0, \ldots, v_s)$ of vertices such that v_{i-1} and v_i are adjacent ($1 \leqslant i \leqslant s$) and $v_{i-1} \neq v_{i+1}$ ($1 \leqslant i \leqslant s - 1$). A group $G \leqslant \mathrm{Aut}(X)$ is *locally s-arc-transitive* on X if for every vertex v_0, the stabilizer of v_0 in G acts transitively on the s-arcs starting at v_0. If in addition G is vertex-transitive then G is *s-arc-transitive.* Otherwise X is clearly bipartite and G acts transitively on each color-class. For $s = 1$, s-arc-transitivity is the same as flag-transitivity.

X is called (locally) s-arc-transitive if the action of Aut(X) is (locally) s-arc-transitive. We shall always assume that X is not a cycle (which is s-arc-transitive for every s).

Having excluded the cycles, local s-arc-transitivity implies large girth: the girth must be $\geqslant 2s - 2$. Hence in a locally s-arc-transitive graph, all s-arcs are paths.

For trivalent s-arc-transitive graphs, Tutte (1947) proved the surprising result that s must be bounded: $s \leqslant 5$ (cf. Biggs 1974, chapter 18). He showed that $s = 5$ is attained by a graph C_8 called an "8-cage", a trivalent graph of girth 8 with 1440 vertices; $\mathrm{Aut}(C_8) \cong \mathrm{Aut}(S_6)$ where S_6 is the symmetric group of degree 6 (cf. Biggs 1974, p. 125). By the covering construction of Theorem 1.6 we infer that there are infinitely many trivalent 5-arc-transitive graphs.

Tutte's result was generalized to *locally* s-arc-transitive graphs in a remarkable self-contained 4-page paper by R.M. Weiss (1976).

Theorem 5.1 (R.M. Weiss). *Let G be a locally s-arc-transitive but not $(s+1)$-arc-transitive group acting on a trivalent graph. Then $s \leqslant 7$ and $s \neq 6$.*

The bound 7 is attained by the 12-*cage* (Tits 1959, Appendix, cf. Benson 1966).

A group G is (locally) s-regular if G is (locally) s-arc-transitive and the stabilizer of each s-arc is the identity. This is a somewhat artificial concept except for degree 3 when it occurs naturally: a trivalent edge-transitive graph is locally s-regular for some s. Let G be a locally s-regular group on a trivalent graph, and G_v a vertex-stabilizer; then $|G_v| = 3 \cdot 2^s$, the number of s-arcs starting at v. Weiss' bound $s \leqslant 7$ thus implies that there is only a finite number of possibilities for the vertex stabilizer in a trivalent edge-transitive graph. These possibilities were classified by Tutte for the flag-transitive (and therefore s-regular) case.

For the edge-transitive (and therefore *locally* s-arc-transitive) case the object to be classified is the pair of vertex-stabilizers of an adjacent pair of vertices together with their intersection $(G_u, G_v, G_u \cap G_v)$. Goldschmidt (1980) classified all these triples and found that there were precisely 15 of them. Goldschmidt's 30-page work is motivated by the examples afforded by the (bipartite) incidence graphs of "buildings" associated with rank-2 BN pairs over GF(2), occurring in the study of certain classes of groups of Lie type. Goldschmidt's "amalgam method" was the starting point of an important new theory (Delgado et al. 1985), used among others for some aspects of "revisionism", a project aiming at a clean and simplified proof of CFSG (Gorenstein et al. 1994).

Tutte's 1947 theorem was extended a third of a century later to s-arc-transitive graphs of arbitrary degree: R. M. Weiss showed, using heavy guns, that $s \leqslant 7$ *holds for s-arc-transitive graphs of arbitrary degree* (Weiss 1981). Noting that the stabilizer G_v of a vertex v in a locally 2-arc-transitive group G acts doubly transitively on the neighbors of v, he was able to invoke the *classification of the doubly transitive permutation groups*, available as a consequence of CFSG (cf. chapter 12). Weiss proves that if $s \geqslant 4$ then the action of G_v on the set $X(v)$ of neighbors is either affine (has an elementary abelian normal subgroup; in particular the degree is a prime power), or it includes the linear fractional group $\mathrm{PSL}(2, p^\alpha)$ as a normal subgroup in its action on the projective line of $|X(v)| = p^\alpha + 1$ points. Here either $s = 4$, or $p \leqslant 3$ and $s \leqslant 2p + 1$.

One of the key ingredients in much of the work on arc-transitive graphs was the following theorem, magically singling out a prime number, characteristic for the graph. The result is due to J. G. Thompson and H. Wielandt and was adapted by Gardiner (1973) in this context (cf. Brouwer et al. 1989, chapter 7.2). For a subset $S \subseteq V(X)$, let $X^d(S)$ denote the set of vertices within distance d from S (so, e.g., $X^0(S) = S$). We use $G_d(S)$ to denote the pointwise stabilizer of $X^d(S)$ in $G \leqslant \mathrm{Aut}(X)$.

Theorem 5.2 (Thompson, Wielandt). *Let $G \leqslant \mathrm{Aut}(X)$ act vertex-transitively on the connected graph X which is not a cycle. Assume that the stabilizer G_v of each vertex v acts as a primitive group on the set of neighbors of v. Then there exists a prime p such that $G_1(e)$ is a p-group (possibly the identity) for every edge e of X.*

Weiss (1979) eliminated the condition of vertex-transitivity and proved that under this weaker assumption (which is implied by *local* 2-arc-transitivity) $G_2(v)$ is a p-group for some vertex v.

No analog of Weiss' $s \leqslant 7$ bound is known for *locally* s-arc-transitive graphs of arbitrary degree. The significance of such an extension would be in its wider applicability which would include incidence graphs of geometries of high symmetry. Such an application of the following partial result of Weiss will be indicated in Theorem 5.5. We should stress that Weiss' proof is *elementary*.

Theorem 5.3 (Weiss 1979). *Let $G \leqslant \mathrm{Aut}(X)$ be a locally s-arc-transitive group acting on the connected graph X of girth g. Assume $s \geqslant 8$ and $g \leqslant 2s + 11$. Then $G_5(S) = 1$ for every arc S of length 14.*

5.2. *Distance-transitive graphs*

This is one of the deepest and most extensively studied areas. We refer to Biggs (1974) for an introduction and to the recent monographs by Brouwer, Cohen and Neumaier (1989) and Bannai and Ito (1984) for technical discussions. The techniques are partly combinatorial and algebraic (adjacency algebras) and apply in greater generality to distance *regular* graphs (cf. chapter 15, section 4); partly group theoretic (both elementary and CFSG-dependent).

First we mention that the *infinite* distance-transitive graphs of finite degree have a very simple structure. For $r, s \geqslant 2$, an *r-tree of s-cliques* is an infinite connected graph all of whose 2-connected blocks are s-cliques and each vertex belongs to exactly r of these cliques.

Theorem 5.4 (Macpherson 1982). *Every infinite distance-transitive graph of finite degree is an r-tree of s-cliques for some $r, s \geqslant 2$.*

Macpherson's proof is based on Dunwoody's theorem on cuts of graphs with more than one end (Theorem 3.37). (Cf. Ivanov's theorem below.) In constrast, a great variety of infinite distance-transitive graphs of infinite degree follows by Fraïssé's theorem (Theorem 5.8) (Cameron, cf. Brouwer et al. 1989, p. 233). Henceforth in this section we assume that our graphs are finite. (Exception: Theorem 5.6.)

Recently, a project aiming at the complete classification of all distance-transitive graphs was drawn up (see the survey by Praeger 1990). There are two phases to this project: to classify vertex-primitive distance-transitive graphs; and to reduce the general case to these. The program of the first phase was layed out by Praeger, Saxl and Yokoyama (1987) who reduced the problem to cases when the automorphism group is either almost simple or affine (has an elementary abelian normal subgroup). As a result of combined efforts of Ivanov, Van Bon, Cohen, Inglis, Liebeck, Praeger, Saxl and others, most of the resulting cases have been settled and this phase now approaches completion (cf. Praeger 1990 for references, and Liebeck, Praeger and Saxl 1987b as an example).

The second phase has not advanced nearly as far but its basic idea is classical.

A graph X of finite diameter d is *antipodal* if being at distance d is an equivalence relation among the vertices of X. Antipodal graphs X of diameter $d \geqslant 2$ are not vertex-primitive since $X^{(d)}$ is disconnected. (In $X^{(k)}$, two points are adjacent if they are at distance k in X.)

The study of distance-transitive graphs can, in a sense, be reduced to the vertex-primitive case, by a result of D. H. Smith and N. J. Martinov which asserts that *a distance-transitive graph of degree $\geqslant 3$ is either primitive, or bipartite, or antipodal.* (Cf. Brouwer et al. 1989, chapter 4.2.) It follows that starting from a distance-transitive graph, two simple operations will eventually lead to a vertex-primitive one. If X is antipodal, we identify antipodes and obtain a distance-transitive graph *covered* by X. If X is bipartite then $X^{(2)}$ has two isomorphic components, both are distance-transitive (X is a *bipartite doubling* of these components). Bipartite doublings have been studied in a number of recent papers (see, e.g., Hemmeter and Woldar 1990). Gardiner's (1974) paper initiated the study of antipodal covers. He showed in particular that the size of the antipodal equivalence class is not greater than the degree. Antipodal coverings of some classes were classified recently (see Liebler 1991, Van Bon and Brouwer 1987).

One of the remarkable general results in this area, predating the classification project indicated, is a classification of distance-transitive graphs by their degree. In 1974, Biggs and Smith (1971) determined all trivalent distance-transitive graphs (there are 12 of them). Smith (1974) went on to determining all tetravalent distance-transitive graphs. Mostly by work of Faradžev et al. (1984) and Ivanov and Ivanov (1988), all distance-transitive graphs of valency $\leqslant 13$ are now known.

Theorem 5.5 (Cameron, Praeger, Saxl and Seitz 1983, Cameron 1982). *There are finitely many distance-transitive graphs of any given degree $d \geqslant 3$.*

For the primitive case, this is immediate from Sims' conjecture (Theorem 1.1, depending on CFSG). Cameron (1982) points out that the general case rapidly follows, observing that from a distance-transitive graph of degree $k \geqslant 3$ the two operations mentioned above (halving, antipodal quotients) lead to a primitive distance-transitive graph of valency $3 \leqslant k' \leqslant k(k-1)$ in at most two steps.

Remarkably, Weiss (1985b) found a proof of Theorem 5.5 avoiding the CFSG reference, based on one of his results on s-arc-transitive graphs (Theorem 5.3), combined with the following powerful elementary result of A. A. Ivanov.

A graph $X = (V, E)$ is *distance-regular* if parameters a_i, b_i, c_i exist such that for each vertex $v \in V$, every vertex at distance i from v has c_i, a_i, and b_i neighbors at distance $i-1$, i, and $i+1$ from v, resp. Distance-transitive graphs are clearly distance-regular. We consider the parameter $t = \sup\{i\colon (a_i, b_i, c_i) = (a_1, b_1, c_1)\}$. (It is clear that $g \leqslant 2t+3$ where g is the girth. If $g \geqslant 4$ then $(a_1, b_1, c_1) = (0, k-1, 1)$ and $2t+2 \leqslant g \leqslant 2t+3$.)

Theorem 5.6 (A. A. Ivanov 1983). *If a distance-regular graph has degree k then its diameter is $d \leqslant t \cdot 4^k$.*

This result is valid for infinite graphs as well, implying that in that case $t = \infty$, hence the graph is an r-tree of s-cliques for some $r, s \geqslant 2$, thus extending Macpherson's theorem to *distance-regular* graphs.

Returning to finite graphs, it is shown in Brouwer et al. (1989, p. 220) via Weiss' proof, that the diameter of a distance-transitive graph of degree k is $d \leqslant (k^6)!4^k$. In reality, $d \leqslant 8$ for $k = 3$, and $d \leqslant 2k-1$ in all known cases for $k \geqslant 4$.

We mention two more parameter bounds. Godsil (1988) proves that *if a distance-regular graph X has an eigenvalue of multiplicity $f \geqslant 3$ then either X is complete multipartite or X has diameter $d \leqslant 3f-4$ and degree $k \leqslant (f-1)(f+2)/2$.* The dodecahedron attains the diameter bound; the icosahedron attains the valency bound.

Using CFSG through the list of doubly transitive groups, Weiss (1985a) classifies the s-arc-transitive graphs of girth $g \leqslant 2s+2$ $(s \geqslant 4)$. (Note that $g \geqslant 2s-2$ always; and $g \leqslant 2s+2$ holds for all distance-transitive graphs.) As a corollary, he finds *all distance-transitive graphs of degree $k \geqslant 3$ and girth $g \geqslant 9$.* In addition to the two largest trivalent distance-transitive graphs (the Biggs–Smith graph on 102 vertices $(g=9)$ and the Foster graph on 90 vertices $(g=10)$, he finds an infinite sequence of graphs with $g=12$, the incidence graphs of the generalized hexagons of associated with the Chevalley groups $G_2(q)$, q a power of 3.

Distance-transitive digraphs are considered by Bannai, Cameron and Kahn (1981).

5.3. *Homogeneity*

In this section we consider a very strong symmetry constraint, the study of which has led to powerful applications of group theory to model theory. A deeper survey is Lachlan (1986); Kantor, Liebeck and Macpherson (1989) is accessible to the reader less versed in model theory.

We shall consider finite and countably infinite graphs, digraphs, and other structures.

A graph X is *homogeneous* if every isomorphism between finite induced subgraphs extends to an automorphism of X. Homogeneous digraphs, hypergraphs, etc. are defined analogously. Clearly, the complement of a homogeneous graph is again homogeneous.

Gardiner (1976) showed that *the only finite homogeneous graphs are $m \cdot K_n$* (the disjoint union of cliques of equal size), *their complements, $L(K_{3,3})$, and the pentagon.* The finite homogeneous tournaments are just the single point and $\vec{C}_3$ (the directed 3-cycle) (Woodrow 1979). The list of finite homogeneous oriented graphs (digraphs with no 2-cycles) is the following: the single point, $\vec{C}_3$, $\vec{C}_3[\overline{K}_m]$ (lexicographic product, section 2), $m \cdot \vec{C}_3$ (m copies of $\vec{C}_3$), $\vec{C}_4$, and finally the Cayley digraph of the quaternion group Q_8 with respect to the generating set $\{i, j, k\}$ in the usual notation (Lachlan).

Cameron 1980a (cf. Cameron, Goethals and Seidel 1978) and Gel'fand (unpublished) strengthened Gardiner's result considerably by relaxing the homogeneity condition. We call the graph X *k-homogeneous* if isomorphisms of subgraphs of $\leqslant k$ vertices extend to automorphisms.

Theorem 5.7 (Cameron, Gel'fand). *If X is a 5-homogeneous finite graph then X appears on Gardiner's list; and therefore X is homogeneous.*

Actually, the result of Cameron and Gel'fand is even more general in that they replace the symmetry condition by a *regularity* condition: X is *k-regular* if any two isomorphic induced subgraphs of $\leqslant k$ vertices have the same number of common neighbors. Observe that "1-regularity" means X is regular; and "2-regularity" means X is strongly regular. These conditions do not imply the presence of any automorphisms and allow a great variety of examples. This fact is in contrast with the situation for $k \geqslant 5$: *If the finite graph X is 5-regular then it appears on Gardiner's list* (Cameron, Gel'fand).

The following generalization allows us to bring graphs of diameter greater than 2 into the picture. Let us call a graph X *metrically k-transitive* if any X-distance preserving map between ordered k-tuples of vertices of X extends to an automorphism of X. Note that for $k = 1$ this is vertex-transitivity, and for $k = 2$ it is distance-transitivity. We also note that the neighborhood of a vertex in a metrically k-transitive graph is $(k-1)$-homogeneous. Building on this fact and on Theorem 5.7, Cameron *classifies all finite metrically 6-transitive graphs.* The connected ones are the complement of $m \cdot K_n$, $K_{n,n}$ with a perfect matching deleted, the cycles, $L(K_{3,3})$, the icosahedron, and the graph $J(6,3)$ on 20 vertices identified with the set of 3-subsets of a 6-set; two vertices are adjacent if the corresponding 3-sets share two elements. It follows that these graphs are automatically metrically k-transitive for every k.

Now we turn to the *countably infinite* (countable for short) case. The best known example is the *Rado graph*, or "generic countable graph", characterized by the following property: given any two disjoint finite subsets A and B of the vertex set, there exists a vertex adjacent to all vertices in A but none in B. This property determines a unique countable graph. The Rado graph contains all finite graphs as induced subgraphs. A countable random graph (each pair is adjacent with probability $\frac{1}{2}$ independently) has probability 1 to be isomorphic to the Rado graph (Erdős and Rényi 1963).

In addition, for every m there exists a unique "generic countable graph without K_m subgraphs", $\mathcal{G}_m$. In this classification, the Rado graph is $\mathcal{G}_\infty$. Lachlan and Woodrow (1980) show that the $\mathcal{G}_m$ $(3 \leqslant m \leqslant \infty)$ and their complements exhaust all nontrivial examples of countable homogeneous graphs; the trivial ones are disjoint unions of cliques of equal size and their complements.

The Rado graph has an obvious tournament analogue, the "generic tournament". Lachlan showed that there are only two other countable homogeneous tournaments: the dense linear order (the order-type of the rationals), and the dense circular order. The latter is defined by a countable dense set on the unit circle with no pairs of antipodal points; edges correspond to clockwise walks along the shorter of the two arcs joining a pair of points.

Homogeneous partial orders were classified by Schmerl (1979); a countable number of them was found. In contrast to these results, Henson (1972) found continuum many nonisomorphic countable homogeneous oriented graphs. Notwithstanding, Cherlin (1987) classified all the homogeneous oriented graphs.

Model theorists' interest in homogeneous structures dates back to a 1954 paper of Fraïssé (1954) linking homogeneity, categoricity, and quantifier elimination.

Let us consider a locally finite "language", i.e. a set L of relation symbols, each associated with a positive integer called the arity such that each arity occurs a finite number of times. An L-structure $\mathcal{M}$ is a set M endowed with a relation of appropriate arity for each symbol in L. (A k-ary relation is a subset of M^k. We allow the case $M = \emptyset$.) Graphs, digraphs correspond to the language of a single binary relation. Every subset of M induces a substructure. (We use the term "substructure" to mean induced substructure.) $\mathcal{M}$ is *homogeneous* if all isomorphisms of finite substructures extend to automorphisms of $\mathcal{M}$. The theory $\mathrm{Th}(\mathcal{M})$ consists of all first-order sentences which are true in $\mathcal{M}$. The theories of homogeneous L-structures are precisely those which permit quantifier elimination (first-order statements of the form $\varphi(u_1,\ldots,u_k)$ depend only on the substructure induced by $u_1,\ldots,u_k$).

Let $\mathcal{F}(\mathcal{M})$ be the class of structures isomorphic to finite substructures of $\mathcal{M}$. A class $\mathcal{C}$ of L-structures is *hereditary* if it is closed under taking substructures. $\mathcal{C}$ has the *amalgamation property* if, whenever $\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2 \in \mathcal{C}$ and $g_i : F_0 \to F_i$ are embeddings (isomorphisms onto substructures) ($i = 1,2$), there exist $\mathcal{F}_3 \in \mathcal{C}$ and embeddings $f_i : F_i \to F_3$ ($i = 1,2$) such that $g_1 f_1 = g_2 f_2$. (Note especially that we allowed the case $F_0 = \emptyset$, thus taking care of what logicians call the "joint embedding property".) An isomorphism-closed class $\mathcal{C}$ of finite L-structures is called an *amalgamation class* if it is hereditary and has the amalgamation property.

Theorem 5.8 (Fraïssé). *If $\mathcal{M}$ is a countable homogeneous L-structure, then $\mathcal{F}(\mathcal{M})$ is an amalgamation class. Conversely, every amalgamation class of finite L-structures is $\mathcal{F}(\mathcal{M})$ for a countable homogeneous L-structure $\mathcal{M}$, unique up to isomorphism.*

The construction of $\mathcal{M}$ in the second statement is a direct limit argument. Since the class of finite graphs without K_m is clearly an amalgamation class, the generic graphs $\mathcal{G}_m$ of the Lachlan–Woodrow theorem are uniquely determined.

A countable structure $\mathcal{M}$ is *$\aleph_0$-categorical* if (up to isomorphism) is the only countable model of its theory. *Every countable homogeneous structure is $\aleph_0$-categorical.* (The converse is false.) $\aleph_0$-categoricity depends solely on $\mathrm{Aut}(\mathcal{M})$. The *$k$-types* of a structure $\mathcal{M}$ are the orbits of $\mathrm{Aut}(\mathcal{M})$ on M^k.

Theorem 5.9 (Ryll-Nardzewski, Engeler, Svenonius). *A countable structure $\mathcal{M}$ is $\aleph_0$-categorical if and only if it has a finite number of k-types for every finite k.*

This result, in a sense, reduces the study of $\aleph_0$-categorical structures to the study of *oligomorphic permutation groups* (groups which have a finite number of orbits on k-sets for every k; see chapter 12, section 9.5, cf. Cameron 1990). Oligomorphic groups are precisely the dense subgroups (w. r. to pointwise convergence) in the automorphism groups of $\aleph_0$-categorical structures over locally finite languages.

$\mathcal{N}$ is a *smooth substructure* of $\mathcal{M}$ if $\mathcal{N}$ is a substructure and (i) all automorphisms of $\mathcal{N}$ extend to $\mathcal{M}$; and (ii) for each k, two k-tuples $u, v \in N^k$ belong to the same k-type of $\mathcal{N}$ if and only if they belong to the same k-type of $\mathcal{M}$.

An $\aleph_0$-categorical L-structure is *smoothly approximable* if it is the union of a chain of finite smooth substructures. The "trivial examples" in the Lachlan-Woodrow theorem, i.e., the disjoint unions of complete graphs and their complements, are smoothly approximable. By Theorem 5.9, the approximating finite structures must have a bounded number of k-types for every fixed k.

In one of the most exciting developments in model theory recently, combined work of Cherlin, Lachlan, Harrington, Kantor, Liebeck, Macpherson, and Hrushovski (Cherlin and Lachlan 1986, Cherlin et al. 1985, Kantor et al. 1989, Cherlin and Hrushovski 1994), heavily relying on CFSG, has led to the *classification of all finite L-structures with a bounded number of 5-types*. (The same magic number 5 as in Theorem 5.7.)

Let $\mathscr{C}(L,k)$ denote the class of L-structures with at most k 5-types. The final result is that $\mathscr{C}(L,k)$ can be decomposed into finitely many classes and each class has a simple dimension theory: a finite number of dimensions is identified, and each first-order statement is equivalent to a Boolean combination of finiteness and exact value statements of each dimension. The dimensions can be varied essentially independently. Dimensions correspond to classes of Lie geometries; the classical examples of the latter are linear and projective spaces over finite fields, possibly with forms (symplectic, orthogonal, unitary), and Grassmannians over disjoint unions of these. Pure sets occur as degenerate examples. The ℓth Grassmannian over the geometry $\mathscr{G}$ is the orbit of an ℓ-dimensional subgeometry of $\mathscr{G}$ under $\mathrm{Aut}(\mathscr{G})$. (When $\mathscr{G}$ is the disjoint union of t pure sets each of size m, the ℓth Grassmannian is the association scheme defined by the natural action of $S_m \wr S_t$ on a set of size $n = \binom{m}{\ell}^t$.)

The proof uses full force of CFSG through the structure theory of primitive permutation groups (O'Nan–Scott theorem, cf. chapter 12), including recent work of Aschbacher and Liebeck on maximal subgroups of classical groups.

A corollary of this theory is that for every finite language L and every fixed k, *membership in $\mathscr{C}(L,k)$ can be tested in polynomial time.*

Another curious corollary is the following. Let us say that the graph X has the *m-extension property* if for any two disjoint subsets A, B of the vertex set there exists a vertex adjacent to all vertices in A but none in B, assuming $|A| + |B| \leqslant m$. (The Rado graph has this property for all m. Almost all graphs on n vertices have the m-extension property for $m = (1-\varepsilon)\log_2 n$; and the Paley graph $P(q,2)$ (section 1.1) for $m = (\frac{1}{2} - \varepsilon)\log_2 q$ (Bollobás 1985, chapter 13.2.))

Corollary 5.10 (Cherlin and Hrushovski 1994). *If for every m, X_m is a finite graph with the m-extension property then the number of orbits of* $\mathrm{Aut}(X_m)$ *on 5-tuples of vertices is unbounded (as $m \to \infty$).*

It would be desirable to see a proof of this result which does not require CFSG.

A final note on higher cardinals: Kierstead and Nyikos (1989) characterize those n-uniform hypergraphs of cardinality κ which have a finite number of isomorphism types of induced subhypergraphs of cardinality λ for some infinite $\lambda < \kappa$.

6. Graph isomorphism

Deciding whether or not two explicitly given finite algebraic or combinatorial structures are isomorphic has been a long-standing unsolved question in the theory of computing. Since all such structures can be canonically encoded in polynomial time by graphs (Hedrlín and Pultr 1966, Miller 1979), it would suffice to solve it for graphs.

From a practical point of view, backtrack algorithms perform quite well. The leader in the trade is McKay's (1987) program "Nauty". However, in spite of considerable effort, the theoretical complexity status of graph isomorphism is still unresolved.

6.1. Complexity theoretic remarks

For basic concepts of computational complexity theory we refer to chapter 29; see also Garey and Johnson (1979).

While "graph isomorphism" (the set of pairs of isomorphic graphs) clearly belongs to NP, it is not known to belong to coNP. In other words, it is not known whether or not for all pairs of nonisomorphic graphs, a short (polynomial length) proof of nonisomorphism exists. It is known, however, that nonisomorphism has bounded round interactive proofs (Goldreich et al. 1986), a fact that puts "nonisomorphism" in the class AM, a randomized extension of NP. This is considered strong theoretical evidence against NP-completeness of "graph isomorphism"; if it were NP-complete, the "polynomial time hierarchy", a hierarchy of complexity classes between P and PSPACE, would collapse. For further references, see Babai and Moran (1988) (cf. chapter 29).

6.2. Algorithmic results: summary of worst case bounds

The best current worst-case bound for a general graph isomorphism algorithm is $\exp\sqrt{cn\log n}$ for n-vertex graphs (Luks and Zemlyachenko, cf. Babai and Luks 1983, and Babai, Kantor and Luks 1983). For some special classes of graphs, substantially better results are available. For groups given by their multiplication tables, and for Steiner triple systems, $n^{O(\log n)}$ isomorphism tests easily follow from the observation that these structures have generating sets of size $\leqslant \log n$. For *planar* graphs, ingenious use of stacks has resulted in a *linear time* isomorphism test (Hopcroft and Tarjan 1972, Hopcroft and Wong 1974). Combinatorial methods in a similar spirit yielded polynomial time isomorphism tests for graphs of bounded genus ($n^{O(g)}$ time for genus $g \geqslant 1$) (Filotti and Mayer 1980). Group theoretic methods led to polynomial time algorithms for graphs with colored vertices and bounded color-classes (isomorphisms preserve colors by definition) (Babai 1979c), for graphs with bounded multiplicity of eigenvalues (Babai, Grigoryev and Mount 1982), and, with considerably deeper use of group theory, for graphs of bounded degree (Luks 1982) ($n^{O(d)}$ time for graphs of degree $\leqslant d$ (Babai and Luks 1983); $O(n^3 \log n)$ time for trivalent graphs, Galil et al. 1987). As a consequence of Luks' methods, isomorphism of block designs (BIBDs) with bounded k and λ can be tested in time $n^{O(\log n)}$ (Babai and Luks 1983) (k is the block size and there are λ blocks common to each pair of vertices); isomorphism of tournaments can be tested in time $n^{O(\log n)}$ (Babai and Luks 1983); and isomorphism of λ-planes (symmetric designs) with bounded λ in $n^{O(\log\log n)}$ time (Babai and Luks 1983).

A common generalization of the polynomial time results for bounded degree and bounded genus was obtained by Miller (1983a,b).

Luks' beautiful paper (Luks 1982) is the single most fundamental reading in the area. It introduces the profound links to group theory to be discussed in section 6.6.

6.3. *Canonical forms*

An algorithmic problem closely related to graph isomorphism is the problem of *complete invariants* and in particular of *canonical forms* of graphs. Let $\mathscr{K}$ denote a class of objects with an equivalence relation to be called "isomorphism". An *invariant* on $\mathscr{K}$ is a mapping f from $\mathscr{K}$ to some class $\mathscr{L}$ of objects such that whenever $X, Y \in \mathscr{K}$ are isomorphic, $f(X) = f(Y)$. We call f a *complete* invariant, if the converse also holds: $f(X) = f(Y)$ implies $X \cong Y$. If, in addition, $\mathscr{L} = \mathscr{K}$ and $f(X) \cong X$ for every $X \in \mathscr{K}$ then the complete invariant f is called a *canonical form* over $\mathscr{K}$; and $f(X)$ the canonical form of X. For graphs, a canonical form f assigns a labeling to the vertices, and this assignment is uniquely defined by f up to automorphisms of X. We call such a labeling canonical, although strictly speaking it is the coset of the automorphism group consisting of all the labelings corresponding to f which is canonical.

Clearly, if a canonical form for a class of objects is available, then isomorphism testing is accomplished by simply comparing the canonical forms. The converse is not known to be true, but in all classes listed above, canonical forms can be obtained within the same time bound as guaranteed for isomorphism testing (cf. Babai and Luks 1983).

An important invariant of graphs is the characteristic polynomial of their adjacency matrix. This invariant fails to be complete (quite badly, cf. Corollary 1.14), as do all other known polynomial time computable invariants.

An example of a canonical form of a graph is the one which produces the lexicographically first adjacency matrix. While this is clearly a complete invariant, unfortunately it is NP-hard to compute (reduction from maximum clique).

6.4. *Combinatorial heuristics: success and failure*

Testing graph isomorphism is easily seen to be equivalent to determining the orbits of the automorphism group of a graph. It is therefore natural to try to find invariant colorings of the vertex set $V(X)$ (i.e. each color class should be a union of orbits of $\mathrm{Aut}(X)$), and refine the color partition in the hope that eventually we obtain the orbit partition. An ordered partition $(C_1, \ldots, C_m)$ of $V(X)$ into invariant color classes C_i can be refined in a simple way: with each vertex $v \in C_i$, we associate the list $(i, \beta_1, \ldots, \beta_m)$, where β_j denotes the number of neighbors of v in C_j. Now order these lists lexicographically; vertices with the same list receive the same color in the new coloring. (The first round colors the vertices by their degree.) Eventually the process stops at a *stable coloring*, characterized by the fact that for every i, j all vertices in C_i have the same number of neighbors in C_j.

Let $\mathscr{T}$ denote the class of graphs which are partitioned by this process into singletons. Clearly, these graph have no automorphisms other than the identity, and the refinement process results in a unique canonical labeling of the graphs belonging to $\mathcal{T}$.

This naive method is highly successful on average: all but an exponentially small fraction of the graphs on n vertices are partitioned into singletons *in the third round* (and thus in linear time) (Babai and Kučera 1979). This is a constructive version of the Erdős-Rényi theorem that all but an exponentially small fraction of the graphs are asymmetric (section 1.6).

Perhaps even more surprising is the result of Kučera (1987) that a modified procedure yields a unique *canonical labeling of almost all trivalent graphs* (and of graphs of bounded degree) *in linear time*. One of the difficulties in handling regular graphs in linear time is how to achieve an initial coloring at all. Kučera achieves this by considering the shortest cycles.

If we allow more time, a simpler way would be to *individualize* a vertex, i.e. to assign a unique color to it, thereby creating a nontrivial initial coloring. Even if subsequent refinements lead to complete partitioning into singletons, we still have to repeat the procedure for every vertex, thereby losing a factor of n in time. One can also individualize a set of k vertices at once (giving each of them distinct colors), thereby increasing the running time by a factor of n^k.

This combination is shown in Babai (1980b, 1981c) to succeed for strongly regular graphs as well as for primitive coherent configurations with $k < 4\sqrt{n}\log n$ (see chapter 41, section 4).

A stronger refinement procedure was proposed in 1968 by Weisfeiler and Lehman (see Weisfeiler 1976): they suggested to color the set of ordered pairs of vertices. Given an ordered partition $V \times V = C_1 \times \cdots \times C_m$, into color classes C_i, we associate with each pair (u, v) of vertices the list $(i, \beta_{jk}\colon 1 \leqslant j, k \leqslant m)$, where $(u, v) \in C_i$ and β_{jk} counts those vertices w with $(u, w) \in C_j$ and $(w, v) \in C_k$. Now again order these lists lexicographically to obtain a refined coloring of $V \times V$. The initial coloring of $V \times V$ uses 3 colors: edges, non-edges, and the diagonal.

The class of graphs for which no refinement is obtained is the *strongly regular graphs*. In general, the stable partitions for the Weisfeiler–Lehman procedure are precisely the *coherent configurations* (chapter 15, section 3).

One can generalize the Weisfeiler–Lehman procedure to partitioning the set V^d of ordered d-tuples in an analogous way. The stable configuration obtained is canonical and the question is, for what d is the resulting partition of the diagonal necessarily the orbit partition of the vertex set. Such a d would yield a canonical form computable in $O(n^{d+1})$ time.

The Cameron–Gel'fand theorem (Theorem 5.7) implies that for $d \geqslant 5$, at least one nontrivial partition occurs in all cases except for the unions of complete graphs of equal size and the complements thereof. The result Babai (1980b) mentioned above implies that $d = O(\sqrt{n}\log n)$ completely succeeds for strongly regular graphs.

Yet a surprising negative result of Cai, Fürer and Immerman (1992) dashed the hopes for a purely combinatorial isomorphism test in moderately exponential

($\exp(n^{1-c})$) time. They construct a *pair of nonisomorphic graphs which force $d = \Omega(n)$ in order for the Weisfeiler–Lehman procedure for d-tuples to distinguish them.*

Their counterexample still leaves ample room for a combination of combinatorial and group theoretic methods to work. Their graphs are partitioned into vertex classes of size 4, and, as mentioned before, the simplest group theoretic method, based on Babai (1979c), yields canonical forms for graphs with bounded color classes in polynomial time.

We should mention that the current best timing for isomorphism testing and canonical forms for general graphs, $\exp(O(\sqrt{n \log n}))$, is obtained by combining Luks' group theoretic method with a combinatorial trick of Zemlyachenko (see Zemlyachenko et al. 1985) (cf. Babai 1981a). Since Zemlyachenko's method does not apply for instance to 3-uniform hypergraphs, the best bound for isomorphism testing within this class is C^n (Luks, cf. Babai and Luks 1983).

6.5. Reductions, isomorphism complete problems, Luks equivalence class

The graph isomorphism problem (ISO for short) is polynomial time equivalent to the isomorphism problem for directed, vertex and edge-colored graphs (isomorphisms preserve colors by definition), and more generally to explicit structures with a set of relations of arbitrary arities. This can be proven by the method of encoding colors into gadgets as in Frucht's theorem (cf. Hedrlín and Pultr 1966, Miller 1979). A number of restricted classes $\mathscr{C}$ are known to be *isomorphism complete*, i.e. ISO can be reduced to isomorphism within $\mathscr{C}$. These include commutative semigroups, k-connected regular bipartite graphs with or without Hamilton cycles, graphs with large girth and chromatic number, etc. Exceptions are those classes which are known to have subexponential ($\exp(n^{o(1)})$) isomorphism tests (groups, latin squares, tournaments, polynomial time testable classes), as well as strongly regular graphs.

The following problems are also known to be equivalent to ISO (see Mathon 1979). Given a graph, determine (i) the orbits of $\mathrm{Aut}(X)$; (ii) generators of $\mathrm{Aut}(X)$; (iii) (Babai and Mathon) the order of $\mathrm{Aut}(X)$.

Observe that (ii), if applied to the union of a pair of isomorphic connected graphs, yields an isomorphism.

Luks found another, related *equivalence class of group theoretic problems*. Let $G, H \leqslant \mathrm{Sym}(\Omega)$ be permutation groups given by a list of generators. The following problems are polynomial time equivalent: (a) find (generators for) $G \cap H$ (*group intersection*); (b) given an element $\sigma \in \mathrm{Sym}(\Omega)$, decide whether or not $G \cap H\sigma = \emptyset$ (*coset intersection*); (c) given a subset $A \subset \Omega$, find the *set-stabilizer* of A in G; (d) given $A \subset \Omega$ and $\sigma \in \mathrm{Sym}(\Omega)$, decide whether the set-stabilizer of A intersects the coset $G\sigma$; (e) given $\sigma, \tau \in \mathrm{Sym}(\Omega)$, decide whether or not σ belongs to the double coset $G\tau H$; (f) given $\tau \in G$, find the centralizer of τ in G; (g) given $\sigma, \tau \in G$, decide whether or not the centralizer of τ in $\mathrm{Sym}(\Omega)$ intersects $G\sigma$.

(Note that if "set-stabilizer" is replaced by "pointwise set stabilizer" in problems (c) and (d), they become polynomial time solvable.)

Proposition 6.1 (Luks). *ISO reduces to coset intersection in polynomial time.*

For simplicity we prove instead, how to reduce the determination of $\mathrm{Aut}(X)$ to group intersection. Let $X = (V, E)$ be a graph and let Ω be the set of unordered pairs from V. Let $G \leqslant \mathrm{Sym}(\Omega)$ denote the induced action of $\mathrm{Sym}(V)$ on pairs; and let $H = \mathrm{Sym}(E) \times \mathrm{Sym}(\Omega \setminus E) \leqslant \mathrm{Sym}(\Omega)$ be the set stabilizer of E in $\mathrm{Sym}(\Omega)$. Then obviously, the induced action of $\mathrm{Aut}(X)$ on Ω is $G \cap H$. □

It is significant that there is strong theoretical evidence suggesting that the decision problems in the Luks equivalence class ((b), (d), (e), (g)) are *not* NP-complete (Goldreich et al. 1986, Babai and Moran 1988). If any of these problems (and therefore each of them) were NP-complete, this would imply the collapse of the "polynomial time hierarchy" in complexity theory, just as NP-completeness of ISO would (cf. section 6.1).

Even more significantly, subcases of ISO can be reduced to polynomial time solvable subcases of *coset intersection*, and thereby they become polynomial time solvable themselves. This is one of the fundamental observations in Luks' (1982) seminal paper.

6.6. *Groups with restricted composition factors*

In this section, we sketch the proof of the main result of Luks (1982).

Theorem 6.2 (Luks). *Isomorphism of graphs of bounded degree can be tested in polynomial time.*

Recall that we used Γ_d to denote the class of groups with a chain of subgroups $G = G_0 \geqslant \cdots \geqslant G_m = 1$ such that $|G_{i-1}:G_i| \leqslant d$. This is the class of groups which occurs as edge-stabilizers in connected graphs of degree $\leqslant (d+1)$ (Theorem 4.4 (c)).

Using the trivial direction of this characterization, Luks reduced isomorphism of graphs of degree $\leqslant (d+1)$ to set stabilizers within a coset $G\sigma$ ($G \leqslant \mathrm{Sym}(\Omega)$, $\sigma \in \mathrm{Sym}(\Omega)$), where $G \in \Gamma_d$ and G is given by a list of generators. Next, he solved the latter problem in polynomial time, inventing a permutation group version of the classical algorithmic technique of "divide and conquer". The idea is to solve the problem one orbit at a time, reducing to a sub-coset in each round. For transitive G, we break G into blocks of imprimitivity; let N be the stabilizer of a system of maximal blocks. Now G/N acts as a primitive group on the blocks. $G\sigma$ is the union of $|G/N|$ cosets of N, and we solve the problem separately inside each coset. Formally, fix $A \subseteq \Omega$, and for any G-invariant set $B \subseteq \Omega$ let $\mathscr{C}(B, G\sigma) = \{\pi \in G\sigma\colon (A \cap B)^\pi = A \cap B\}$. This set is either empty or a coset of a subgroup of G. The identity $\mathscr{C}(B_1 \cup B_2, G\sigma) = \mathscr{C}(B_1, \mathscr{C}(B_2, G\sigma))$ is used to reduce to the transitive case. For $H \leqslant G$ we have $G = \bigcup_i H\tau_i$ and thus $\mathscr{C}(B, G\sigma) = \bigcup \mathscr{C}(B, H\tau_i\sigma)$; setting $H = N$, this can be used to reduce the imprimitive case to the intransitive case.

The algorithm runs in polynomial time because of the following result. It is easy to see that Γ_d can be characterized as the class of groups of which each composition factor is a subgroup of the symmetric group S_d.

Theorem 6.3 (Babai, Cameron and Pálfy 1982). *Let $G \leqslant S_n$ be a primitive group of degree n and assume $G \in \Gamma_d$. Then $|G| \leqslant n^{cd}$ where c is an absolute constant. More generally, if all alternating composition factors of G have bounded orders and all classical groups among the composition factors of G have bounded dimensions then $|G| \leqslant n^C$ for some constant C depending only on the bounds in the condition.*

Note that in particular, *primitive solvable groups* have order $\leqslant n^c$, where $c = 3.24399...$ (Pálfy 1982, Wolf 1982).

Turning back to Luks' algorithm, Theorem 6.3 guarantees that $|G/N|$ is polynomially bounded, allowing a recurrence in timing with polynomially bounded solution, completing the proof of Theorem 6.2. (We note that Theorem 6.3 was not available to Luks at the time; instead, in the difficult affine case, he used the second reduction step above with H a Sylow p-subgroup which he showed had polynomially bounded index.) □

Isomorphism of tournaments can be decided in $n^{O(\log n)}$ time (Babai and Luks 1983). This algorithm uses the Pálfy–Wolf bound on primitive solvable groups (above) (and the Feit–Thompson theorem through the solvability of the automorphism groups of tournaments).

6.7. *Basic permutation group algorithms*

We assume in this section that a permutation group $G \leqslant \mathrm{Sym}(\Omega)$ is given by a set S of s generators; $|\Omega| = n$. Some of the basic algorithmic problems to solve are testing *membership* in G of a given $\sigma \in \mathrm{Sym}(\Omega)$; determining the *order* of G; constructing the *normal closure* of a subgroup (also given by a list of generators). Once these are solved, solvability and nilpotence of G are easily decided. In his pioneering work in computational group theory, Sims (1970, 1971, 1978) constructed algorithms for these problems which ran fast in practice and were later asymptotically analysed to run in polynomial time in the worst case (see below).

Theory and practice diverge in the areas of more advanced problems, including determining the *center*, the *composition factors*, the *Sylow subgroups*. All these problems are now solvable in polynomial time. The elegant construction of a composition chain and the composition factors (Luks 1987) uses the O'Nan–Scott Theorem (chapter 12) and requires CFSG (the classification of finite simple groups) through Schreier's hypothesis (the outer automorphism group of a simple group is solvable). Beals (1993b) has recently found an elementary algorithm for composition factors. Kantor's (1985a,b) construction of the Sylow subgroups starts with finding a composition chain via Luks (1987) and rests on detailed knowledge of CFSG and a case-by-case study of the classical groups. Luks' (1987) algorithm to find the center is elementary.

Many other important problems are *not* known to be solvable in polynomial time, and in fact often they are at least as hard in general as *graph isomorphism* (centralizers, intersections, cf. section 6.5). Particularly efficient backtrack procedures have recently been found and implemented by Leon (1991), using partitioning heuristics (cf. section 6.4). Such procedures are often used even for problems solvable in polynomial time (e.g., finding the center by repeated application of a backtrack routine for centralizers), showing a discrepancy between theoretical and practical measures of efficiency.

For the rest of this section we return to the complexity analysis of the basic problems. Given a chain $G = G_0 \geqslant G_1 \geqslant \cdots \geqslant G_m = 1$ of subgroups, a *strong generating set* (SGS) with respect to this chain is a set $T \subseteq G$ such that $\langle T \cap G_i \rangle = G_i$ for every i. This concept was introduced by Sims (1970) (with respect to the stabilizer chain) as the fundamental data structure for permutation group algorithms. (Recent algorithms often operate on different chains of subgroups; however, it is possible to switch efficiently from any SGS to one in Sims' sense, Cooperman et al. 1990.) Given an SGS, the problems of membership and order can be solved easily, a presentation (in terms of generators and relations) can be deduced, and slight variations of the SGS methods yield normal closures as well. Variants of Sims' method have been shown to run in polynomial time ($\mathrm{O}(n^6 + sn^2)$ (Furst, Hopcroft and Luks 1980) and $\mathrm{O}(n^5 + sn^2)$ Knuth 1991, Jerrum 1986). These elementary algorithms require $\Omega(n^5)$ even on *average* on large classes of examples (Knuth 1991).

Better asymptotic bounds have been obtained using heavy guns. For two functions f, g let us write $f(n) = \mathrm{O}^{\sim}(g(n))$ if for sufficiently large n, $f(n) \leqslant g(n) \log^c n$ for some constant c. With this notation, the best current deterministic asymptotic worst case bound is $\mathrm{O}^{\sim}(sn^3)$ (Babai, Luks and Seress 1993). This bound depends on CFSG primarily through estimates of the orders of primitive permutation groups (chapter 12, Theorem 5.8, cf. Cameron 1981). With randomization we can do considerably better and have an entirely elementary $\mathrm{O}^{\sim}(n^3 + sn)$ Monte Carlo algorithm to construct an SGS (Babai, Cooperman, Finkelstein, Luks and Seress 1991). (Being Monte Carlo, the algorithm does not guarantee to construct an SGS but it does so with arbitrarily large probability.) The algorithm includes a particularly efficient *normal closure* routine, running in $\mathrm{O}^{\sim}(n^2 + sn)$. The basic technique of the algorithm generalizes the following observation: *Let $g_1, \ldots, g_k \in G$ generate G and let H be a proper subgroup of G. Then the probability that $h \notin H$ for a* random subproduct $h = g_1^{\varepsilon_1} \cdots g_k^{\varepsilon_k}$ *is* $\geqslant \frac{1}{2}$. (The $\varepsilon_i \in \{0, 1\}$ are selected by independent unbiased coin-flips.)

A *base* of G is a set $B \subseteq \Omega$ such that the pointwise stabilizer of B in G is the identity. Let $\mu(G)$ be the minimum size of a base. The case of small $\mu(G)$ is of particular interest. For instance, if G is simple non-alternating then $\mu(G) = \mathrm{O}(\log n)$. It is easy to see that $2^{\mu(G)} \leqslant |G| \leqslant n^{\mu(G)}$. Let us say that a class $\mathscr{G}$ of groups has *small bases* if $\mu(G) = (\log n)^{\mathrm{O}(1)}$ for $G \in \mathscr{G}$. Sims style algorithms run in $\mathrm{O}^{\sim}(sn^2)$ on groups with a small base. Using new combinatorial techniques, elementary Monte Carlo algorithms have been found which construct an SGS in *nearly linear*, $\mathrm{O}^{\sim}(sn)$ time for small base groups (Babai, Cooperman, Finkelstein and Seress 1991). The speedup relies on methods capable of handling chains of certain *subsets* of G which are not subgroups; the subgroup structure of small base groups tends to be too coarse to allow nearly linear time. The key new ingredients are an efficient implementation of Sims' *Schreier vector data structure* to store coset representatives in a *shallow tree* (depth guaranteed to be $\leqslant \log |G|$) via an algorithmic version of the Reachability theorem (Theorem 6.4); and the use of the Local expansion property (Theorem 3.41) to rapidly locate new elements if the current partial SGS misses a substantial portion of G.

Finding *domains of imprimitivity* seems indispensable when delving deeper into the group structure. Atkinson's (1975) algorithm finds them in quadratic time. For small base groups, Beals (1993a) improved this to nearly linear time, and in a *tour de force,* used this with Seress to find a composition series in nearly linear time (Beals and Seress 1992).

A final note on *parallelization.* An NC *algorithm* uses $n^{O(1)}$ parallel processors and extremely short, $(\log n)^{O(1)}$ time, where n is the length of the input. (So none of the processors has time to read any substantial portion of the input; cf. chapter 29.) Radical departure from the classical methods has allowed the design of an NC *algorithm to construct an SGS* and solve some of the basic problems in NC, including membership, order, normal closures, solvability, center, composition factors (Babai et al. 1987). Again, the algorithm uses CFSG mainly through Theorem 5.8 of chapter 12, and also requires Luks' composition factors algorithm. The algorithm digs deeply into the normal structure of G. Even the rudimentary task of membership testing requires determining the composition series first.

6.8. *Complexity of related problems*

Problems related to *graph isomorphism* ("ISO" for short) and *permutation group membership* fall into a variety of complexity classes. Groups, semigroups will be given by a list of generators, unless otherwise stated.

A surprising result of Lubiw (1981) asserts that the following problem is NP-complete: *Does a given permutation group have a fixed-point-free element*? Even the case when G is an elementary abelian 2-group is NP-complete. Lalonde (1981) used this to show the following problem NP-complete: *Does a given bipartite graph have an automorphism of order* 2 *interchanging the two color classes*? In contrast, if we omit the "order 2" restriction, the problem becomes *isomorphism complete* (equivalent to ISO). The original (equivalent) statement of Lalonde's theorem is this: *The star system problem is* NP-*complete.* The "star system problem" has a family $\mathcal{F}$ of n subsets of an n-set V for input and asks if there exists a graph $X = (V, E)$ such that $\mathcal{F} = \{X(v)\colon v \in V\}$ is the family of vertex neighborhoods in X.

Isomorphism of groups of order n, given by their Cayley tables, can be decided in time $n^{\log_2 n+O(1)}$ because the groups are generated by $\leqslant \log_2 n$ elements and any mapping of the generators can be extended to a homomorphism in at most one way. This argument generalizes to quasigroups which in turn include Steiner triple systems.

To decide *isomorphism of permutation groups* is at least as hard as ISO (Babai, Kannan and Luks 1994). On the other hand this problem is in NP for the following simple reason: Let $G = \langle S\rangle \leqslant \mathrm{Sym}(A)$ and $H = \langle T\rangle \leqslant \mathrm{Sym}(B)$ be permutation groups and $f\colon S \to \mathrm{Sym}(B)$ a map. Then f extends to an isomorphism of G onto H if and only if the following two polynomial time testable conditions hold: (i) H is generated by the f-image of S; (ii) the orders of G, H, and the group $\langle (s,f(s)) \colon s \in S \rangle$ agree. On the other hand, isomorphism of permutation groups also belongs to the class coAM (Babai, Kannan and Luks 1994) (cf. section 6.1) and is therefore unlikely to be NP-complete.

If $G, H, K \leqslant \mathrm{Sym}(A)$ and $\sigma \in \mathrm{Sym}(A)$ then the *double coset membership* problem "$\sigma \in GH$?" belongs to the Luks equivalence class (is equivalent to coset intersection) (section 6.5). On the other hand, the question "$\sigma \in GHK$?" is NP-complete (Luks).

The membership problem for *semigroups of transformations* of a finite set is PSPACE-complete (Kozen 1977).

The membership problem for $d \times d$ *integral matrices* is *undecidable* already for $d = 4$. This is immediate from the following result of Mihailova (1958): The membership problem is undecidable for subgroups of $F_2 \times F_2$, where F_2 is the free group of rank 2. However, finiteness of an integral matrix group (or a matrix group over an algebraic number field) can be decided in polynomial time (Babai et al. 1993), and if the group is finite, the usual basic questions (order, center, composition chain, Sylow subgroups) can be answered in Las Vegas polynomial time (Beals and Babai 1993). (A Las Vegas algorithm uses randomization but never outputs a wrong answer.)

For finite groups, the membership problem is in NP under quite general conditions. A *black box group* is, informally, a group whose elements are encoded by (0,1)-strings of uniform length, and the group operations are performed by a "black box". (As all our groups, a black box group is given by a list of generators.) Then membership is in NP, relative to the black box. In particular, membership in matrix groups over finite fields is in NP. This is immediate from the following combinatorial result. A *straight line program* reaching a group element $g \in G$ from a set S of generators of G is a sequence $g_1, \ldots, g_m$ of elements of G such that $g_m = g$, and for each i, either $g_i \in S$, or $g_i = g_j^{-1}$, or $g_i = g_j g_k$ for some $j, k < i$. The *cost* of such a program is the number of inversions and multiplications (the calls to S are free). The *straight line cost* of $g \in G$ (relative to S) is the minimum cost of straight line programs reaching g from S.

Theorem 6.4 (Reachability theorem, Babai and Szemerédi 1984). *Given any set S of generators of a group G of order n, the straight line cost of any $g \in G$ is less than* $(1 + \log_2 n)^2$.

We conjecture that membership in matrix groups also belongs to coNP. The proof of this statement and the stronger statement that the *order* of a matrix group over a finite field belongs to NP (i.e. the correct order has polynomial time verifiable certificates) depends, in essence, on the following conjecture.

Short presentation conjecture. Every group of order n has a presentation (in terms of generators and relations) of length $(\log n)^{O(1)}$.

(The *length* of a presentation is the total number of characters required to write down the presentation.) It follows from Theorem 6.4 that it suffices to prove this conjecture for simple groups. All cases have been confirmed with the exception of the rank 1 simple groups of twisted Lie type (unitary, Suzuki, Ree) (Babai, Goodman, Kantor, Luks and Pálfy 1994).

None of the problems mentioned in this section, with the possible exception of isomorphism of groups given by a Cayley table, is expected to have polynomial time solution. In particular, the membership problem for 1×1 matrix groups is a close relative of the *discrete logarithm* problem (given $a, b \in \mathrm{GF}(q)$, find an integer x such that $a^x = b$ or decide that no such x exists) which is not expected to be solvable in polynomial time (cf. Adleman and Demarrais 1993).

Modulo this obstacle, however, a great deal of stucture can be found in matrix groups and even in black box groups (Beals and Babai 1993).

7. The reconstruction problem

All graphs in this section are finite unless otherwise stated.

In the Introduction to this chapter we gave a general definition of reconstructibility; and discussed a number of instances. Examples include Whitney's theorems on the reconstructibility of graphs from their line graphs (with known exceptions) (section 1.2), of 3-connected graphs from their cycle matroids (cf. chapter 11, section 7), and from many other functions of graphs (the area of graph equations comes under this heading, see Cvetković 1979). The unsettled status of the Graph isomorphism problem is related to the non-reconstructibility from any of the known polynomial time computable invariants.

While reconstruction problems (solved and unsolved) seem to pop up in nearly every topic considered, the term "Reconstruction problem" has been reserved for a single notorious member of this species in graph theory: the Kelly–Ulam reconstruction conjecture. It is this problem to which this brief last section is devoted. For more information and references we refer to the surveys mentioned in the preface to this chapter.

7.1. Vertex reconstruction

With every graph $X = (V, E)$ we associate the multiset $D^{\mathrm{v}}(X)$ of isomorphism types of its one-vertex-deleted subgraphs, i.e. the isomorphism type of $X \setminus v$ for each $v \in V$. We call $D^{\mathrm{v}}(X)$ the *deck of* 1-*vertex-deleted subgraphs*. Analogously one can define the multiset $D^{\mathrm{e}}(X)$, the deck of 1-edge-deleted subgraphs, and more generally, $D^{\mathrm{v}}_k(X)$ and $D^{\mathrm{e}}_k(X)$, the decks of k-vertex-deleted (k-edge-deleted, resp.) subgraphs.

The graph X is *vertex-reconstructible* (or simply reconstructible) if it is determined (up to isomorphism) by $D^{\mathrm{v}}(X)$. Edge-reconstructibility is defined analogously. More generally we say that the graph invariant $f(X)$ (cf. section 6.3) is vertex-reconstructible if $f(X)$ is determined by $D^{\mathrm{v}}(X)$. The *Reconstruction conjecture* says that *all finite graphs with* $\geqslant 3$ *vertices are reconstructible* (Kelly and Ulam in 1942).

The answer to the analogous question for directed graphs is negative: an infinite family of pairs of non-isomorphic tournaments with identical decks has been found by Stockmeyer (1977).

It is known that *almost every graph is vertex-reconstructible* (Erdős). Indeed, this is an immediate consequence of the fact that almost every graph X has the following property: no pair of two-vertex-deleted subgraphs of X are isomorphic. This argument generalizes to smaller subgraphs: almost all graphs are reconstructible from their k-vertex-deleted subgraphs for all $k < c \log n$ for some constant $c > 0$.

Some concrete classes of graphs are also known to be reconstructible. These include disconnected graphs, trees (Kelly in 1957), and some families of tree-like graphs. In particular, all graphs with $\leqslant n$ edges are reconstructible.

Among the reconstructible invariants, one should mention the degree sequence and a refinement of this: the sequence of degree sequences of the neighborhoods of the vertices (Nash-Williams 1978). Applying powerful counting techniques to reconstruction theory, Tutte (1979) has shown important polynomials associated with graphs to be reconstructible: the characteristic polynomial, the chromatic polynomial, and generalizations of these.

The Reconstruction conjecture is false for infinite graphs (even for forests) but no counterexamples are known to the following variant, *Halin's conjecture:* If two (finite or infinite) graphs with at least 3 vertices have the same deck of vertex-deleted subgraphs, then each is isomorphic to a subgraph of the other.

7.2. *Edge reconstruction*

It is known that a vertex-reconstructible graph with at least 4 edges is also edge-reconstructible (Greenwell 1971). In addition, however, large classes of graphs are known to be edge-reconstructible for which vertex-reconstructibility is open. The first result in this direction was Lovász's (1972b) who proved that if a graph has more edges than its complement then it is edge-reconstructible. Lovász's proof used a clever inclusion–exclusion argument which was the basis of rapid further improvements. Müller (1977) showed that graphs with m edges and n vertices are edge-reconstructible unless $2^{m-1} \leqslant n!$, which means $m \leqslant n \cdot \log_2 n$. Nash-Williams (1978) modified Müller's proof and obtained the following lemma, from which Müller's bound is immediate.

Lemma 7.1 (Nash-Williams). *Suppose that the graph $X = (V, E)$ is not edge-reconstructible. Then for every subset $A \subseteq E$ such that $|A \setminus E|$ is even, there exists a permutation $\sigma \in \mathrm{Sym}(V)$ such that $E \cap E^{\sigma} = A$.*

Lovász observed that this lemma has the following immediate consequence.

Corollary 7.2. *If $X = (V, E)$ is not edge-reconstructible then for every $T \subseteq E$,*

$$|\sigma \in \mathrm{Sym}(V)\colon\ T^{\sigma} \subseteq E| \geqslant 2^{|E|-|T|-1}.$$

Pyber (1990) used this to derive that *all Hamiltonian graphs are edge-reconstructible*, with possibly a finite set of exceptions. Indeed, by Corollary 7.2, a nonreconstructible Hamiltonian graph with n vertices and m edges would have at least $2^{m-n-2}/n$ Hamilton cycles. But this is too much: Pyber proves that no graph has more than c^{m-n} Hamilton cycles, where $c = 1.977$. □

The arguments used in the proofs of Lovász, Müller, Nash-Williams lend themselves to a much more general treatment. The following framework was introduced by Mnukhin (1987).

Let $G \leqslant \mathrm{Sym}(\Omega)$ be a permutation group acting on the set Ω. We say that two subsets $\Delta_1, \Delta_2 \subseteq \Omega$ are G-isomorphic if $\Delta_1^\sigma = \Delta_2$ for some $\sigma \in G$. For any subset $\Gamma \subseteq \Omega$ let Γ^G be the G-orbit of Γ, i.e. the set of subsets of Ω, G-isomorphic to Γ.

For $\Delta \subseteq \Omega$ let the k-deleted deck $D_k(\Delta)$ be the multiset of G-isomorphism classes of the $(|\Delta| - k)$-element subsets of Δ. The set Δ is k-reconstructible if it is determined (up to G-isomorphism) by its k-deleted deck $D_k(\Delta)$.

In particular, taking Ω to be the set of $\binom{n}{2}$ pairs of elements of V and $G \cong \mathrm{Sym}(V)$ be the induced action of $\mathrm{Sym}(V)$ on Ω, the concept of G-isomorphism of subsets of Ω becomes the ordinary isomorphism of graphs on the vertex set V; and k-reconstructibility turns into the concept of reconstructibility from the deck of k-*edge-deleted* subgraphs.

Generalizing Müller's theorem Mnukhin proves that if $\Delta \subset \Omega$ is not 1-reconstructible then $2^{|\Delta|-1} \leqslant |G|$.

Below we indicate a *linear algebra* approach introduced by Godsil et al. (1987) to extend Müller's result to k-reconstructibility for $k \geqslant 2$. Their technique is easily adapted to Mnukhin's situation.

Recall that a hypergraph $\mathcal{F} \subseteq 2^\Omega$ is m-uniform if $|E| = m$ for each $E \in \mathcal{F}$.

Definition. The *Vapnik–Chervonenkis dimension* or VC dimension of a hypergraph $\mathcal{F} \subseteq 2^\Omega$ is the greatest integer t for which there exists a subset $A \subseteq \Omega$ with $|A| = t$ such that every subset $B \subseteq A$ occurs as $B = A \cap E$ for some $E \in \mathcal{F}$.

For $0 \leqslant s \leqslant n$ the s-inclusion matrix $I(\mathcal{F}, s)$ of a hypergraph $\mathcal{F} \subseteq 2^\Omega$ has rows indexed by the members $F \in \mathcal{F}$, columns indexed by subsets $A \subseteq \Omega$ with $|A| = s$, and entry 1 if $A \subseteq F$ and 0 otherwise. The s^*-inclusion matrix $I^*(\mathcal{F}, s)$ has all the columns of the t-inclusion matrices for $t = 0, 1, 2, \ldots, s$.

We say that $\mathcal{F}$ is *s-independent* if the rows of the s-inclusion matrix are linearly independent (i.e. $I(\mathcal{F}, s)$ has full row-rank), and it is s^*-independent if the rows of $I^*(\mathcal{F}, s)$ are linearly independent. Clearly s-independence implies s^*-independence, and for uniform hypergraphs, the converse also holds (Frankl and Wilson 1981).

Theorem 7.3 (Frankl and Pach 1983). *If $\mathcal{F}$ is s^*-dependent, then its VC dimension is at least $s + 1$.*

The proof follows from the proof of Corollary 4.2 in chapter 31. For a theory of the inclusion matrices, including this result, see Babai and Frankl 1992.

The main lemma of Godsil et al. (1987) follows.

Lemma 7.4. *If Δ_1 and Δ_2 have the same k-deleted deck $D_k(\Delta_i)$ but are not G-isomorphic, then the m-uniform set-system $\mathcal{F} = \Delta_1^G \cup \Delta_2^G$ is $(m-k)$-dependent (where $m = |\Delta_i|$).*

Proof. We prove the dependence of the rows of $I(\mathcal{F}, m-k)$ by explicitly giving coefficients $c(E)$ $(E \in \mathcal{F})$ for a linear relation among them. For $i = 1,2$ let $\alpha_i = |G_{\{\Delta_i\}}|$ (the size of the set-stabilizer of Δ_i). If $E \in \Delta_1^G$ let $c(E) = \alpha_1$, and if $E \in \Delta_2^G$ let $c(E) = -\alpha_2$. To check that this linear combination of the rows is a zero row, consider a column indexed by a set $T \subseteq \Omega$ with $|T| = m - k$. The column has zeros except where $T \subseteq E$. So the entry for this column in the indicated linear combination of the rows will be α_1 times the number of $E \in \Delta_1^G$ with $T \subseteq E$, minus α_2 times the number of $E \in \Delta_2^G$ with $T \subseteq E$. This is the number of $\sigma \in G$ for which $T \subseteq \Delta_1^\sigma$ minus the number of $\sigma \in G$ for which $T \subseteq \Delta_2^\sigma$. But this difference is zero because for every set T of size $m-k$, the number of $\sigma \in G$ for which $T^\sigma \subseteq \Delta_i$ is independent of i. □

Using this lemma and Theorem 7.3 we infer the following generalization of Müller's inequality.

Theorem 7.5 (Godsil, Krasikov and Roditty 1987). *If $\Delta \subseteq \Omega$ is not k-reconstructible, then $2^{|\Delta|-k} \leqslant |G|$.*

Proof. Combining the foregoing results we obtain that for $\mathcal{F}$ as before, the VC-dimension of $\mathcal{F}$ is $\geqslant m-k+1$. Hence $|\mathcal{F}| \geqslant 2^{m-k+1}$, while clearly $|\mathcal{F}| \leqslant 2|G|$. □

In particular we obtain that if a graph with n vertices and m edges is not k-reconstructible then $2^{m-k} \leqslant n!$, or $m \leqslant k + n \log_2 n$.

For $k = 1$ we also recover Lovász's corollary to the Nash-Williams lemma (slightly improved).

Theorem 7.6. *If $\Delta \subseteq \Omega$ is not 1-reconstructible, then for every $\Gamma \subseteq \Delta$,*

$$|\{\sigma \in G\colon \Gamma^\sigma \subseteq \Delta\}| \geqslant 2^{|\Delta|-|\Gamma|} - 1.$$

Proof. Let $\mathcal{F}$ be as before (now $k = 1$). Since its VC dimension is $\geqslant m - k + 1 = m$, there is a set $A \subseteq \Omega$ with $|A| = m$ of which every subset is its intersection with some $E \in \mathcal{F}$. In particular, $A \in \mathcal{F}$. Now take any proper subset Γ of Δ. Since Δ and Δ_2 have the same 1-deleted deck, we also have $\Gamma^\sigma \subseteq \Delta_2$ for some $\sigma \in G$, hence we have $\Gamma^\tau \subseteq A$ for some $\tau \in G$ (since $A \in \mathcal{F} = \Delta^G \cup \Delta_2^G$). And $|\{\sigma \in G\colon \Gamma^\sigma \subseteq \Delta\}| = |\{\sigma \in G\colon \Gamma^\tau \subseteq \Delta^\sigma\}|$. But this latter is at least the number of proper subsets of A which contain Γ^τ, because each of those is $A \cap E$ for some $E \in \mathcal{F}$ (hence for some $E = \Delta^\sigma$ since the proper subsets are in the 1-deleted deck which Δ and Δ_2 share). The latter number is $2^{m-|\Gamma|} - 1$. □

References

Abbott, H.L.

[1972] A note on Ramsey's theorem, *Canad. Math. Bull.* **15**, 9–10.

Adleman, L.M., and J. Demarrais

[1993] A subexponential algorithm for discrete logarithms over all finite fields, *Math. Comput.* **61**, 1–15.

Aldous, D.

[1983] Random walks on finite groups and rapidly mixing Markov chains, in: *Séminaire de Probabilités XVII, Lecture Notes in Mathematics*, Vol. 986 (Springer, Berlin) pp. 243–297.

[1987] On the Markov chain simulation method for uniform combinatorial distributions and simulated annealing, *Probab. Eng. Inform. Sci.* **1**, 33–46.

Aldous, D., and P. Diaconis

[1987] Strong uniform times and finite random walks, *Adv. in Appl. Math.* **8**, 69–97.

Alekseiev, V.E.

[1974] On the number of Steiner triple systems, *Math. Notes* **15**, 461–464.

Alon, N.

[1986] Eigenvalues and expanders, *Combinatorica* **6**, 83–96.

Alspach, B.

[1979] Hamiltonian cycles in vertex-transitive graphs of order $2p$, in: *Proc. 10th South-Eastern Conf., Boca Raton, FL, Congress. Numerantium* **XXIII**, 131–139.

Annexstein, F., and M. Baumslag

[1989] *Limitations on Constructing Expanders with Cayley Graphs*, Manuscript.

Aschbacher, M.

[1976] A homomorphism theorem for finite graphs, *Proc. Amer. Math. Soc.* **54**, 468–471.

[1980] *The Finite Simple Groups and Their Classification, Yale Mathematical Monographs*, Vol. 7 (Yale University Press, New Haven, CT).

Aschbacher, M., et al.

[1985] eds., *Proc. Rutgers Group Theory Year, 1983–1984* (Cambridge University Press, Cambridge).

Atkinson, M.D.

[1975] An algorithm for finding the blocks, *Math. Comp.* **29**, 911–913.

Aurenhammer, F., J. Hagauer and W. Imrich

[1992] Cartesian graph factorization at logarithmic cost per edge, *Comput. Complexity* **2**, 331–349.

Babai, L.

[1973] Groups of graphs on given surfaces, *Acta Math. Acad. Sci. Hungar.* **24**, 215–221.

[1974a] Automorphism groups of graphs and edge-contraction, *Discrete Math.* **8**, 13–20.

[1974b] On the minimum order of graphs with given group, *Canad. Math. Bull.* **17**, 467–470. MR 53#10641.

[1975] Automorphism groups of planar graphs II, in: *Infinite and Finite Sets, Proc. Conf. Keszthely*, eds. A. Hajnal, R. Rado and V.T. Sós, *Colloq. Math. Soc. János Bolyai* **1**, 29–84.

[1977a] Some applications of graph contractions, *J. Graph Theory* **1**, 125–130.

[1977b] Symmetry groups of vertex transitive polytopes, *Geom. Dedicata* **6**, 331–338.

[1978a] Chromatic number and subgraphs of Cayley graphs, in: *Theory and Applications of Graphs, Lecture Notes in Mathematics*, Vol. 642, eds. Y. Alavi and D.R. Lick (Springer, Berlin) pp. 10–22.

[1978b] Embedding graphs in Cayley graphs, in: *Combinatories et Theorie des Graphes, Proc. Conf. Paris-Orsay, 1976*, eds. J.-C. Bermond et al. (CNRS, Paris) pp. 13–15.

[1978c] Infinite digraphs with given regular automorphism groups, *J. Combin. Theory B* **25**, 26–46. MR 58#16380.

[1979a] Almost all Steiner triple systems are asymmetric, in: *Topics in Steiner Systems*, eds. C.C. Lindner and A. Rosa, *Ann. Discrete Math.* 7, 37–39.

[1979b] Long cycles in vertex-transitive graphs, *J. Graph Theory* **3**, 301–304. MR 80m:05059.

[1979c] *Monte Carlo Algorithms in Graph Isomorphism Testing*, Tech. Rep. 79-10 (Dép. Math. et Stat., University de Montréal).

[1979d] Spectra of Cayley graphs, *J. Combin. Theory B* **29**, 180–189.

[1980a] Finite digraphs with given regular automorphism groups, *Period. Math. Hungar.* **11**, 257–270.

[1980b] On the complexity of canonical labelling of strongly regular graphs, *SIAM J. Comput.* **9**, 212–216.

[1981a] Moderately exponential bound for graph isomorphism, in: *Fundamentals of Computation Theory, Lecture Notes in Mathematics*, Vol. 117, ed. F. Gécseg (Springer, Berlin) pp. 34–50.

[1981b] On the abstract group of automorphisms, in: *Combinatorics, London Mathematical Society Lecture Note Series*, Vol. 52 (Cambridge University Press, London) pp. 1–40.

[1981c] On the order of uniprimitive permutation groups, *Ann. of Math.* **113**, 553–568.

[1985] Arc transitive covering digraphs and their eigenvalues, *J. Graph Theory* **8**, 363–370.

[1989] The probability of generating the symmetric group, *J. Combin. Theory A* **52**, 148–153.

[1991a] Computational complexity in finite groups, in: *Proc. Int. Congress of Mathematicians, Kyoto 1990* (Springer, Tokyo) pp. 1479–1489.

[1991b] Local expansion of vertex-transitive graphs and random generation in finite groups, in: *Proc. 23rd ACM Symposium on Theory of Computing* (ACM, New York) pp. 164–174.

[1991c] Vertex-transitive graphs and vertex-transitive maps, *J. Graph Theory* **15**, 587–627.

[1994] Vertex-transitive graphs, excluded minors, and hyperbolic geometry, in preparation.

Babai, L., R. Beals and D. Rockmore

[1993] Deciding finiteness of matrix groups in deterministic polynomial time, in: *Proc. ISSAC'93, Kiev* (ACM Press) pp. 117–126.

Babai, L., and P.J. Cameron

[1994a] Automorphism groups of switching classes of tournaments, to appear.

[1994b] Most primitive groups are automorphism groups of edge-transitive hypergraphs, to appear.

Babai, L., P.J. Cameron and P.P. Palfy

[1982] On the orders of primitive groups with restricted non-abelian composition factors, *J. Algebra* **79**, 161–168.

Babai, L., G. Cooperman, L. Finkelstein, E.M. Luks and Á. Seress

[1991] Fast Monte Carlo algorithms for permutation groups, in: *Proc. 23rd ACM Symposium on Theory of Computing* (ACM, New York) pp. 90–100.

Babai, L., G. Cooperman, L. Finkelstein and Á. Seress

[1991] Nearly linear time algorithms for permutation groups with a small base, in: *Proc. ISSAC'91 (Int. Symp. on Symbolic and Algebraic Computation), Bonn* (ACM Press) pp. 200–209.

Babai, L., and P. Erdős

[1982] Representation of group elements as short products, in: *Theory and Practice of Combinatorics*, eds. J. Turgeon, A. Rosa and G. Sabidussi, *Ann. Discrete Math.* **12**, 21–26.

Babai, L., and P. Frankl

[1992] *Linear Algebra Methods in Combinatorics* (The University of Chicago). Preliminary version 2.

Babai, L., and C.D. Godsil

[1982] On the automorphism groups of almost all Cayley graphs, *European J. Combin.* **3**, 9–15.

Babai, L., A. Goodman, W.M. Kantor, E.M. Luks and P.P. Pálfy

[1994] *Short Presentations for Finite Groups*, Manuscript.

Babai, L., and A.J. Goodman

[1991] Subdirectly reducible groups and edge-minimal graphs with given automorphism group, *J. London Math. Soc.* **47**, 417–432.

[1993] On the abstract group of automorphisms, in: *Coding Theory, Design Theory, Group Theory, Proc. Marshall Hall Conf.*, eds. D. Jungnickel and S.A. Vanstone (Wiley, New York) pp. 121–143.

Babai, L., A.J. Goodman and L. Lovász

[1991] Graphs with given automorphism group and few edge orbits, *European J. Combin.* **12**, 185–203.

Babai, L., D.Yu. Grigoryev and D.M. Mount

[1982] Isomorphism of graphs with bounded eigenvalue multiplicity, in: *Proc. 14th ACM Symposium on Theory of Computing* (ACM, New York) pp. 310–324.

Babai, L., and G. Hetyei

[1992] On the diameter of random Cayley graphs of the symmetric group, *Combin. Probab. Comput.* **1**, 201–208.

Babai, L., G. Hetyei, W.M. Kantor, A. Lubotzky and Á. Seress

[1990] On the diameter of finite groups, in: *Proc. 31st IEEE Symp. Found. of Computer Science*, pp. 857–865.

Babai, L., S. Kannan and E.M. Luks

[1994] Bounded round interactive proofs for nonisomorphism of permutation groups, in preparation.

Babai, L., W.M. Kantor and A. Lubotzky

[1989] Small diameter Cayley graphs for finite simple groups, *European J. Combin.* **10**, 507–522.

Babai, L., W.M. Kantor and E.M. Luks

[1983] Computational complexity and the classification of finite simple groups, in: *Proc. 24th IEEE Symp. on Foundations of Computer Science*, pp. 507–522.

Babai, L., and L. Kučera

[1979] Canonical labeling of graphs in linear average time, in: *Proc. 20th IEEE Symp. on Foundations of Computer Science*, pp. 39–46.

Babai, L., and L. Lovász

[1973] Permutation groups and almost regular graphs, *Studia Sci. Math. Hungar.* **8**, 145–150.

Babai, L., and E.M. Luks

[1983] Canonical labeling of graphs, in: *Proc. 15th ACM Symposium on Theory of Computing* (ACM, New York) pp. 171–183.

Babai, L., E.M. Luks and Á. Seress

[1987] Permutation groups in NC, in: *Proc. 19th ACM Symposium on Theory of Computing* (ACM, New York) pp. 409–420.

[1993] Computing composition series in primitive groups, in: *Groups and Computation*, eds. L. Finkelstein and W.M. Kantor, *DIMACS Ser. Discrete Math. Theor. Comput. Sci.* **11**, 1–16.

Babai, L., and S. Moran

[1988] Arthur-Merlin games: a randomized proof system, and a hierarchy of complexity classes, *J. Comput. Syst. Sci.* **36**, 254–276.

Babai, L., and A. Pultr

[1980] Endomorphism monoids and topological subgraphs of graphs, *J. Combin. Theory B* **38**, 278–283. MR 82c:05052.

Babai, L., and Á. Seress

[1988] On the diameter of Cayley graphs of the symmetric group, *J. Combin. Theory A* **49**, 175–179.

Babai, L., and V.T. Sós

[1985] Sidon sets in groups and induced subgraphs of Cayley graphs, *European J. Combin.*, pp. 101–114.

Babai, L., and M. Szegedy

[1992] Local expansion of symmetrical graphs, *Combin. Probab. Comput.* **1**, 1–11.

Babai, L., and E. Szemerédi

[1984] On the complexity of matrix group problems, in: *Proc. 24th IEEE Symp. Found. of Computer Science*, pp. 229–240.

Babai, L., and M. Watkins

[1980] Connectivity of infinite graphs having a transitive torsion group action, *Arch. Math.* **34**, 90–96.

Baer, R.

[1947] Direct decompositions, *Trans. Amer. Math. Soc.* **62**, 62–98.

Bannai, E., P.J. Cameron and J. Kahn

[1981] Nonexistence of certain distance-transitive digraphs, *J. Combin. Theory B* **31**, 105–110.

Bannai, E., and T. Ito

[1984] *Algebraic Combinatorics, I* (Benjamin/Cummings, Menlo Park, CA).

Bass, H.

[1972] The degree of polynomial growth of finitely generated nilpotent groups, *Proc. London Math. Soc.* **25**, 603–614.

Beals, R.

[1993a] Computing blocks of imprimitivity for small base groups in nearly linear time, in: *Groups and*

Computation, eds. L. Finkelstein and W.M. Kantor, *DIMACS Ser. Discrete Math. Theor. Comput. Sci.* **11**, 17–26.

[1993b] An elementary algorithm for computing the composition factors of a permutation group, in: *Proc. ISSAC'93, Kiev* (ACM Press, New York) pp. 127–134.

Beals, R., and L. Babai

[1993] Las Vegas algorithms for matrix groups, in: *Proc. 34th IEEE Symp. on Foundations of Computer Science*, pp. 427–436.

Beals, R., and Á. Seress

[1992] Structure forest and composition factors in nearly linear time, in: *Proc. 24th ACM Symp. on Theory of Computing* (ACM, New York) pp. 116–125.

Bednarek, A.R.

[1985] Whitney's theorem for infinite graphs, *Discrete Math.* **56**, 83–85. MR 87b:05065.

Benson, C.T.

[1966] Minimal regular graphs of girth eight and twelve, *Canad. J. Math.* **18**, 1091–1094.

Berge, C.

[1972] Une condition pour qu'un hypergraphe soit fortement isomorphe à un hypergraphe complet ou multiparti, *C.R. Acad. Sci. Paris* **274**, 1783–1786.

Biggs, N.L.

[1972] The symplectic representation of map automorphisms, *Bull. London Math. Soc.* **4**, 303–306.

[1973] Three remarkable graphs, *Canad. J. Math.* **25**, 397–411.

[1974] *Algebraic Graph Theory, Cambridge Tracts in Mathematics*, Vol. 67 (Cambridge University Press, Cambridge).

Biggs, N.L., and D.H. Smith

[1971] On trivalent graphs, *Bull. London Math. Soc.* **3**, 155–158.

Birkhoff, G.

[1945] Sobre los grupos di automorfismos, *Rev. Union Math. Argentina* **11**, 155–157.

Blass, A., F. Harary and Z. Miller

[1980] Which trees are link graphs? *J. Combin. Theory B* **29**, 277–292.

Bollobás, B.

[1979] *Graph Theory* (Springer, New York).

[1982] The asymptotic number of unlabelled regular graphs, *J. London Math. Soc.* **26**, 201–206.

[1985] *Random Graphs* (Academic Press, London).

Bondy, J.A.

[1991] A graph reconstructor's manual, in: *Surveys in Combinatorics, Proc. 13th British Combinatorial Conf., London Mathematical Society Lecture Notes Series*, Vol. 166, ed. A.D. Keedwell (Cambridge University Press, Cambridge).

Bondy, J.A., and R.L. Hemminger

[1977] Graph reconstruction – a survey, *J. Graph Theory* **1**, 227–268.

Bondy, J.A., and M. Simonovits

[1980] Longest cycles in 3-connected cubic graphs, *Canad. J. Math.* **32**, 987–992.

Bouwer, I.Z.

[1969] Section graphs for finite permutation groups, *J. Combin. Theory* **6**, 378–386.

[1972] On edge but not vertex transitive regular graphs, *J. Combin. Theory B* **12**, 32–40.

Brouwer, A.E., A.M. Cohen and A. Neumaier

[1989] *Distance-Regular Graphs* (Springer, Berlin).

Brown, M., and R. Connelly

[1975] On graphs with a constant link, II, *Discrete Math.* **11**, 199–232.

Bruen, A., and B. Levinger

[1973] A theorem on permutations of a finite field, *Canad. J. Math.* **25**, 1060–1065.

Bulitko, V.K.

[1972] On the problem of the finiteness of a graph with given vertex neighborhoods, in: *General Systems Theory* (in Russian) (Akad. Nauk Ukrain. SSR Inst. Kibernet.) pp. 76–83.

Burnside, W.

[1911] *Theory of Groups of Finite Order* (Cambridge University Press, Cambridge).

Cai, J., M. Fürer and N. Immerman

[1992] An optimal lower bound on the number of variables for graph identification, *Combinatorica* **12**, 389–410.

Cameron, P.J.

[1979] unpublished.

[1980a] 6-transitive graphs, *J. Combin. Theory B* **28**, 168–179.

[1980b] On graphs with given automorphism group, *European J. Combin. 1, 91–96.*

[1981] Finite permutation groups and finite simple groups, *Bull. London Math. Soc.* **13**, 1–22.

[1982] There are only finitely many distance transitive graphs of given valency greater than two, *Combinatorica* **2**, 9–13. MR 83k:05050.

[1983] Automorphism groups of graphs, in: *Selected Topics in Graph Theory*, Vol. 2, eds. R.J. Wilson and L.W. Beineke (Academic Press, New York) pp. 89–127, MR 86i:05079.

[1990] *Oligomorphic Permutation Groups, London Mathematical Society Lecture Note Series*, Vol. 152 (Cambridge University Press, Cambridge).

Cameron, P.J., J.-M. Goethals and J.J. Seidel

[1978] Strongly regular graphs having strongly regular subconstituents, *J. Algebra* **55**, 257–280.

Cameron, P.J., C.E. Praeger, J. Saxl and G.M. Seitz

[1983] On the Sims conjecture and distance transitive graphs, *Bull. London Math. Soc.* **15**, 499–506.

Carlitz, L.

[1960] A theorem on permutations in a finite field, *Proc. Amer. Math. Soc.* **11**, 456–459.

Celler, F., C.R. Leedham-Green, S.H. Murray, A.C. Niemeyer and E.A. O'Brien

[1995] *Generating Random Elements of a Finite Group*, Res. Rep. MRR-013-95 (Australian National University, Canberra).

Chang, C.C.

[1961] Ordinal factorization of finite relations, *Trans. Amer. Math. Soc.* **101**, 259293.

Cherlin, G., and E. Hrushovski

[1994] Finite structures with few types, in preparation.

Cherlin, G.L.

[1987] Homogeneous directed graphs. The imprimitive case, in: *Logic Colloquium '85* (Elsevier, Amsterdam) pp. 67–88.

Cherlin, G.L., L. Harrington and A.H. Lachlan

[1985] $\aleph_0$-categorical, $\aleph_0$-stable structures, *Ann. Pure Appl. Logic* **28**, 103–135.

Cherlin, G.L., and A.H. Lachlan

[1986] Stable finitely homogeneous structures, *Trans. Amer. Math. Soc.* **296**, 815–850.

Chillag, D.

[1988] Generalized circulant and class functions of finite groups II, *Lin. Algebra Appl.* **108**, 199–212.

Cohen, D.E.

[1972] *Groups of Cohomological Dimension One, Lecture Notes in Mathematics*, Vol. 254 (Springer, Berlin).

Conder, M.D.E.

[1980] Generators of the alternating and symmetric groups, *J. London Math. Soc.* **2**, 75–86.

Conway, J.A., N.J.A. Sloane and A.R. Wilks

[1989] Gray codes for reflection groups, *Graphs Combin.* **5**, 315–325.

Cooperman, G., L. Finkelstein and N. Sarawagi

[1990] A random base change algorithm for permutation groups, in: *Proc. Int. Symp. on Symbolic and Algebraic Comp. (ISSAC'90)*, eds. S. Watanabe and M. Nagata (ACM, New York) pp. 161–168.

Coxeter, H.S.M.

[1961] *Introduction to Geometry* (Wiley, New York).

Coxeter, H.S.M., and W.O.J. Moser

[1972] *Generators and Relations for Discrete Groups*, 3rd Ed. (Springer, Berlin).

Cvetković, D.M., M. Doob and H. Sachs
[1980] *Spectra of Graphs* (Academic Press, New York).

Cvetković, D.M., and S.K. Simić
[1979] A bibliography of graph equations, *J. Graph Theory* **3**, 311–324.

de Groot, J.
[1958] Automorphism groups of rings, in: *Int. Congr. of Mathematicians, Edinburgh*, p. 18. Abstract.
[1959] Groups represented by homeomorphism groups I, *Math. Ann.* **138**, 80–102.

Delgado, A., D. Goldschmidt and B. Stellmacher
[1985] *Groups and Graphs, New Results and Methods* (Birkhäuser, Basel).

Delsarte, Ph.
[1973] An algebraic approach to the association schemes of coding theory, *Philips Res. Rep. Suppl.* **10**. MR 52#5187.

Deza, M.
[1973] Une propriété extrémale des plans projectifs finis dans une classe de codes equidistants, *Discrete Math.* **6**, 343–352.

Diaconis, P.
[1988] *Group Representations in Probability and Statistics* (Institute of Mathematical Statistics, Hayward, CA).
[1989] Patterned matrices, in: *Matrix Theory and Application, Proc. Symp. on Applied Mathematics, Phoenix, AZ, 1989* (Oxford University Press, Oxford).

Dicks, W., and M.J. Dunwoody
[1989] *Groups acting on Graphs. Cambridge Studies in Advanced Mathematics*, Vol. 17 (Cambridge University Press, Cambridge).

Dixon, J.D.
[1969] The probability of generating the symmetric group, *Math. Z.* **110**, 199–205.

Doyle, P.G., and J.L. Snell
[1984] *Random Walks and Electric Networks* (Mathematical Association of America, New York).

Driscoll, J.R., and M.L. Furst
[1987] Computing short generator sequences, *Inform. and Comput.* **72**, 117–132.

Dunwoody, M.J.
[1982] Cutting up graphs, *Combinatorica* **2**, 13–25.

Dürnberger, E.
[1985] Every connected Cayley graph of a group with prime order commutator group has a Hamilton cycle, in: *Cycles in Graphs, Burnaby, B.C., 1982, North-Holland Mathematics Studies*, Vol. 115 (North-Holland, Amsterdam) pp. 75–80. MR 87f:05082.

Eilenberg, S.
[1934] Sur les transformations périodiques de la surface de sphére, *Fund. Math.* **22**, 28–41.

Ellingham, M.N.
[1988] Recent progress in edge-reconstruction, *Congress. Numerantium* **62**, 3–20.

Elspas, B., and J. Turner
[1970] Graphs with circulant adjacency-matrices, *J. Combin. Theory* **9**, 297–307. MR 42#7540.

Erdős, P., and A. Rényi
[1963] Assymetric graphs, *Acta Math. Acad. Sci. Hungar.* **14**, 295–315. MR 27#3538.
[1965] Probabilistic methods in group theory, *J. Anal. Math.* **14**, 127–138.

Erdős, Péter L., and Z. Füredi
[1980] On automorphisms of line-graphs, *European J. Combin.* **1**, 341–345.

Even, S., and O. Goldreich
[1981] The minimum length generator sequence is NP-hard, *J. Algorithms* **2**, 311–313.

Faradžev, I.A., A.A. Ivanov and A.V. Ivanov
[1984] Distance transitive graphs of valency 5, 6, and 7, *Zh. Vychisl. Mat. i Mat. Fiz.* **24**, 1704–1718 [1986, *European J. Combin.* 7, 303–319].

Feder, T.

[1992] Product graph representations, *J. Graph Theory* **16**, 467–488.

Feigenbaum, J., and A.A. Schäffer

[1992] Finding the prime factors of strong direct product graphs in polynomial time, *Discrete Math.* **109**, 77–102.

Feit, W.

[1989] Some finite groups with nontrivial centers which are Galois groups, in: *Group Theory*, eds. K.N. Cheng and Y.K. Leong (Walter de Gruyter, Berlin) pp. 87–109.

Fejes Tóth, L.

[1965] *Reguläre Figuren* (Akadémiai Kiadó, Budapest).

Feller, W.

[1968] *An Introduction to Probability Theory and Its Applications I*, 3rd Ed. (Wiley, New York).

Fiat, A., S. Moses, A. Shamir, I. Shimsoni and G. Tardos

[1989] Planning and learning in permutation groups, in: *Proc. 30th IEEE Symp. on Foundations of Computer Science*, pp. 274–279.

Filotti, I., and J. Mayer

[1980] Polynomial-time algorithm for determining the isomorphism of graphs of fixed genus, in: *Proc. 12th ACM Symp. on Theory of Computing* (ACM, New York) pp. 236–243.

Fischer, B.

[1971] Finite groups generated by 3-transpositions, I, *Invent. Math.* **13**, 232–246.

Folkman, J.

[1967] Regular line-symmetric graphs, *J. Combin. Theory* **3**, 215–232.

Fournier, J.C.

[1974] Une condition pour qu'un hypergraphe, ou son complementaire, soit fortement isomorphe à un hypergraphe complet, in: *Hypergraph Seminar, Springer Lecture Notes in Mathematics*, Vol. 411 (Springer, Berlin) pp. 98–98.

Fraïssé, R.

[1954] Sur l'extension aux relations de quelques proprictes des ordres, *Ann. Sci. École Norm. Sup.* **71**, 361–388.

Frankl, P., and J. Pach

[1983] On the number of sets in a null t-design, *European J. Combin.* **4**, 205–236.

Frankl, P., and R.M. Wilson

[1981] Intersection theorems with geometric consequences, *Combinatorica* **1**, 357–368.

Freudenthal, H.

[1945] Über die Enden diskreter Räume und Gruppen, *Comment. Math. Helv.* **17**, 1–38.

Fried, E., and J. Kollár

[1978] Automorphism groups of algebraic number fields, *Math. Z.* **163**, 121–123.

[1981] Automorphism groups of fields, in: *Universal Algebra, Proc. Conf. Esztergom, 1977*, eds. B. Csákány, E. Fried and E.T. Schmidt, *Colloq. Math. Soc. János Bolyai* **24**.

Frucht, R.

[1938] Herstellung von Graphen mit vorgegebener abstrakter Gruppe, *Compos. Math.* **6**, 239–250.

[1949] Graphs of degree 3 with given abstract group, *Canad. J. Math.* **1**, 365–378.

[1952] A one-regular graph of degree three, *Canad. J. Math* **4**, 240–247.

Furst, M.L., J. Hopcroft and E.M. Luks

[1980] Polynomial time algorithms for permutation groups, in: *Proc. 21st IEEE Symp. on Foundations of Computer Science*, pp. 36–41.

Galil, Z., C.M. Hoffmann, E.M. Luks, C.P. Schnorr and A. Weber

[1987] An $O(n^3 \log n)$ deterministic and an $O(n^3)$ Las Vegas isomorphism test for trivalent graphs, *J. Assoc. Comput. Mach.* **34**, 513–531.

Gallai, T.

[1971] Transitiv orientierbare Graphen, *Acta Math. Sci. Hungar.* **22**, 51–G3.

Gardiner, A.

[1973] Arc-transivity in graphs I, *Quart. J. Math. Oxford (2)* **24**, 399–407.

[1974] Antipodal covering graphs, *J. Combin. Theory B* **16**, 255–273.

[1976] Homogeneous graphs, *J. Combin. Theory B* **20**, 94–102. MR 54#7316.

Garey, M.R., and D.S. Johnson

[1979] *Computers and Intractibility, A Guide to the Theory of NP-Completeness* (Freeman, San Francisco).

Gel'fand, I.M., and V.A. Ponomarev

[1970] Problems of linear algebra and classification of quadruples of subspaces in a finite-dimensional vector space, in: *Hilbert Space operators, Colloq. Math. Soc. János Bolyai* 5, 163–237.

Godsil, C.D.

[1978] Graphs, groups and polytopes, in: *Combinatorial Mathematics, Lecture Notes in Mathematics*, Vol. 686 (Springer, Berlin) pp. 157–164. MR 80m:05052.

[1979] *Graphs with regular groups*, Ph.D. Thesis (University of Melbourne).

[1980a] More odd graph theory, *Discrete Math.* **32**, 205–207.

[1980b] Neighborhoods of transitive graphs and GRR's, *J. Combin. Theory B* **29**, 116–140.

[1981a] GRR's for non-solvable groups, in: *Algebraic Methods in Graph Theory, Proc. Conf. Szeged, 1978*, eds. L. Lovász et al., *Coll. Math. Soc. János Bolyai* **25**, 221–239.

[1981b] On the full automorphism group of a graph, *Combinatorica* **1**, 243–256.

[1982] Eigenvalues of graphs and digraphs, *Lin. Algebra Appl.* **46**, 43–50.

[1988] Bounding the diameter of distance-regular graphs, *Combinatorica* **8**, 333–343.

Godsil, C.D., and W. Imrich

[1987] Embedding graphs in Cayley graphs, *Graphs Combin.* **3**, 39–43.

Godsil, C.D., L. Krasikov and Y. Roditty

[1987] Reconstructing graphs from their k-edge-deleted subgraphs, *J. Combin. Theory B* **43**, 360–363.

Goldreich, O., S. Micali and A. Wigderson

[1986] Proofs that yield nothing but their validity and a methodology of cryptographic protocol design, in: *Proc. 27th IEEE Symp. on Foundations of Computer Science*, pp. 168–195.

Goldschmidt, D.M.

[1980] Automorphisms of trivalent graphs, *Ann. of Math.* **111**, 377–406.

Goodman, A.J.

[1993] The edge-orbit conjecture of Babai, *J. Combin. Theory B* **57**, 26–35.

Gorenstein, D., R. Lyons and R. Solomon

[1994] *The Classification of Finite Simple Groups, A.M.S. Math. Surveys and Monographs*, Vol. 40 (AMS, Providence, RI).

Graham, R.L., and P.M. Winkler

[1985] On isometric embeddings of graphs, *Trans. Amer Math. Soc.* **288**, 527–536.

Greenwell, D.L.

[1971] Reconstructing graphs, *Proc. Amer. Math. Soc.* **30**, 431–433.

Greenwell, D.L., and L. Lovász

[1974] Applications of product colouring, *Acta Math. Acad. Sci. Hungar.* **25**, 335–340.

Grigorchuk, R.I.

[1983] On Milnor's problem of group growth, *Soviet Math. Dokl.* **28**, 23–26.

Gromov, M.

[1981] Groups of polynomial growth and expanding maps, *Publ. Math. IHES* **53**, 53–73.

Gross, J.L., and T.W. Tucker

[1987] *Topological Graph Theory* (Wiley, New York).

Grünbaum, B., and G.C. Shephard

[1981] The geometry of planar graphs, in: *Combinatorics, Proc. 8th British Combinatorics Conf.*, ed. H.N.V. Temperley (Cambridge University Press, Cambridge) pp. 124–150.

Hajnal, A.

[1985] The chromatic number of the product of two $\aleph_0$-chromatic graphs can be countable, *Combinatorica* **5**, 137–139.

Halin, R.

[1964] Über unendliche Wege in Graphen, *Math. Ann.* **157**, 125–137. MR 30#578.

Hamermesh, M.

[1962] *Group Theory and its Application to Physical Problems* (Addison-Wesley, Reading, MA).

Hamidoune, Y.O.

[1981] An application of connectivity theory in graphs to factorizations of elements in groups, *European J. Combin.* **2**, 349–355.

[1990] On some graphic aspects of addition theorems, in: *Topics in Combinatorics and Graph Theory*, eds. R. Bodendiek and R. Henn (Physica Verlag, Heidelberg) pp. 349–355.

Harary, F.

[1969] *Graph Theory* (Addison-Wesley, Reading, MA).

Hedrlin, Z., and J. Lambek

[1969] How comprehensive is the category of semigroups? *J. Algebra* **11**, 195–212.

Hedrlin, Z., and A. Pultr

[1965] Symmetric relations (undirected graphs) with given semigroups, *Monatsh. Math.* **69**, 318–322.

[1966] On full embeddings of categories of algebras, *Illinois J. Math.* **10**, 392–406.

Hell, P.

[1978] Graphs with given neighbourhoods I, in: *Proc. Colloq. Int. CNRS, Orsay, 1976* (CNRS, Paris) pp. 219–223.

Hemmeter, J., and A. Woldar

[1990] On the maximal cliques of the quadratic forms graph in even characteristic, *European J. Combin.* **11**, 119–126.

Henson, C.W.

[1972] Countable homogeneous relational structures and $\aleph_0$-categorical theories, *J. Symbolic Logic* **37**, 494–500.

Hering, C.

[1967] Eine Bemerkung über Automorphismengruppen von endlichen projektiven Ebenen und Möbiusebenen, *Arch. Math.* **18**, 107–110.

Hetzel, D.

[1976] *Über reguläre graphische Darstellungen von auflösbaren Gruppen*, Ph.D. Thesis (Technische Universität Berlin). Diplomarbeit.

Holt, D.F.

[1981] A graph which is edge transitive but not arc transitive, *J. Graph Theory* **5**, 201–204.

Hopcroft, J.E., and R.E. Tarjan

[1972] Isomorphism of planar graphs, in: *Complexity of Computer Computations*, eds. R.M. Miller and J.W. Thatcher (Plenum Press, New York) pp. 131–152.

[1973] Dividing a graph into triconnected components, *SIAM J. Comput.* **2**, 135–158.

Hopcroft, J.E., and J.K. Wong

[1974] Linear time algorithm for isomorphism of planar graphs, in: *Proc. 6th ACM Symposium on Theory of Computing* (ACM, New York) pp. 172–184.

Hopf, H.

[1944] Enden offener Räume und unendliche diskontinuierliche Gruppen, *Comment. Math. Helv.* **16**, 81–100.

Hrushovski, E.

[1992] Extending partial isomorphisms of graphs, *Combinatorica* **12**, 411–416.

Huppert, B.

[1957] Zweifach transitive auflösbare Permutationsgruppen, *Math. Z.* **68**, 126–150.

[1967] *Endliche Gruppen I* (Springer, Berlin).

Hurwitz, A.

[1893] Über algebraische Gebilde mit eindeutigen Transformationen in sich, *Math. Ann.* **41**, 403–442.

Imrich, W.

[1971] Assoziative Produkte von Graphen, *Sitzungsber. II, Österr. Akad. Wiss., Math. Naturw. Kl.* **180**, 203–239. MR 47#4863.

[1977] Subgroup theorems and graphs, in: *Combinatorial Mathematics V, Springer Lecture Notes in Mathematics*, Vol. 622 (Springer, Berlin) pp. 1–27.

[1989] Embedding graphs into cartesian products, in: *Proc. 1st China–USA Int. Conf. on Graph Theory and its Applications*, eds. M.F. Capobianco et al., *Ann. N.Y. Acad. Sci.* **576**, 266–274.

[1993] *Graph products, a survey*, unpublished.

Ivanov, A.A.

[1983] Bounding the diameter of a distance regular graph, *Soviet Math. Dokl.* **28**, 149–152.

Ivanov, A.A., and A.V. Ivanov

[1988] Distance-transitive graphs of valency $8 \leqslant k \leqslant 13$, in: *Algebraic, Extremal and Metric Combinatorics* (Cambridge University Press, Cambridge) pp. 112–145.

Jackson, B.

[1986] Longest cycles in 3-connected cubic graphs, *J. Combin. Theory B* **41**, 17–26.

Jerrum, M.

[1985] The complexity of finding minimum-length generator sequences, *Theor. Comput. Sci.* **36**, 265–289.

Jerrum, M.R.

[1986] A compact representation for permutation groups, *J. Algorithms* **7**, 60–78.

Johnson, S.M.

[1963] Generation of permutations by adjacent transpositions, *Math. Comput.* **17**, 282–285.

Jones, G.A., and D. Singerman

[1978] Theory of maps on orientable surfaces, *Proc. London Math. Soc.* **37**, 273–307.

Jónsson, B.

[1981] Arithmetic of ordered sets, in: *Ordered Sets, NATO ASI Ser. C*, Vol. 83, ed. I. Rival (Reidel, Dordrecht) pp. 3–41.

Jordan, C.

[1869] Sur les assemblages de lignes, *J. Reine Angew. Math.* **70**, 185–190.

[1895] Nouvelles recherches sur la limite de transitivité des groupes qui ne contiennent pas le groupe alterné, *J. Math. Paris* **1**, 35 60.

Jungerman, M., and A.T. White

[1980] On the genus of finite abelian groups, *European J. Combin.* **1**, 243–251.

Kantor, W.M.

[1969] Automorphism groups of designs, *Math. Z.* **109**, 246–252. MR 43#71.

[1972] k-homogeneous groups, *Math. Z.* **124**, 261–265.

[1985a] Polynomial-time algorithms for finding elements of prime order and Sylow subgroups, *J. Algorithms* **6**, 478–514.

[1985b] Sylow's theorem in polynomial time, *JCSS* **30**, 359–394.

[1992] Some large trivalent graphs having small diameters, *Discrete Appl. Math.* **37/38**, 353–357.

Kantor, W.M., M.W. Liebeck and H.D. Macpherson

[1989] $\aleph_0$-categorical structures smoothly approximable by finite substructures, *Proc. London Math. Soc.* **59**, 439–463.

Keating, K., and D. Witte

[1985] On Hamilton cycles in Cayley graphs in groups with cyclic-commutator subgroup, in: *Cycles in Graphs, Burnaby, B.C., 1982*, eds. B.R. Alspach and C.D. Godsil, *North-Holland Mathematics Studies*, Vol. 115 (North-Holland, Amsterdam). MR 87f:05082.

Kerékjártó, B. v.

[1921] Über die periodischen Transformationen der Kreisscheibe und der Kugelfläche, *Math. Ann.* **80**, 36–38.

Kesten, H.

[1959] Symmetric random walks on groups, *Trans. Amer. Math. Soc.* **92**, 336–354.

Kierstead, H.A., and P.J. Nyikos

[1989] Hypergraphs with finitely many isomorphism subtypes, *Trans. Amer. Math. Soc.* **312**, 699–718.

Klin, M.H.

[1981] On edge but not vertex transitive graphs, in: *Algebraic Methods in Graph Theory, Proc. Szeged, 1978*, eds. L. Lovász and V.T. Sós, *Colloq. Math. Soc. János Bolyai* **25**, 405 434.

Klin, M.H., M.E. Muzychuk and L.A. Faradžev

[1991] Cellular rings and groups of automorphisms of graphs, in: *Investigations in the Algebraic Theory of Combinatorial Objects* (Kluwer, Dordrecht).

Knuth, D.E.

[1991] Notes on efficient representation of perm groups, *Combinatorica* **11**, 57–68, Preliminary version circulated since 1981.

[1994] The sandwich theorem, *Electron. J. Combin.* **1**. 48pp.

Kompel'macher, V.L., and V.A. Liskovets

[1975] Sequential generation of permutations by means of a basis of transpositions (in Russian), *Kibernetika* **3**, 17–21.

König, D.

[1936] *Theorie der endlichen und unendlichen Graphen* (Akademische Verlaggesellschaft Geest u. Portig, Leipzig).

Kozen, D.

[1977] Lower bounds for natural proof systems, in: *Proc. 18th ACM Symposium on Theory of Computing* (ACM, New York) pp. 254–266.

Kučera, L.

[1987] Canonical labeling of regular graphs in linear average time, in: *Proc. 28th Annu. IEEE Symp. on Theory of Computing*, pp. 271–279.

Lachlan, A.H.

[1986] Homogeneous structures, in: *Proc. Int. Congr. of Mathematicians, Berkely, CA*, ed. A.M. Gleason (AMS, Providence, RI) pp. 314–321.

Lachlan, A.H., and R.E. Woodrow

[1980] Countable ultrahomogeneous undirected graphs, *Trans. Amer. Math. Soc.* **262**, 51–94.

Lalonde, F.

[1981] Le probleme d'etoiles pour graphes est NP-complet, *Discrete Math.* **33**, 271–280.

Leighton, F.T.

[1992] *Introduction to Parallel Algorithms and Architectures: Arrays, Trees, Hypercubes* (Morgan Kaufman, San Mateo, CA).

Leon, J.S.

[1991] Permutation group algorithms based on partitions I: Theory and algorithms, *J. Symbolic Comput.* **12**, 533–583.

Liebeck, M.W.

[1983] On graphs whose full automorphism group is an alternating group or a finite classical group, *Proc. London Math. Soc.* **47**, 337–362.

iebeck, M.W., C.E. Praeger and J. Saxl

[1987a] A classification of the maximal subgroups of the finite alternating and symmetric groups, *J. Algebra* **111**, 365–383.

[1987b] Distance transitive graphs with symmetric or alternating automorphism group, *Bull. Aust. Math. Soc.* **35**, 1–25.

[1988] On the 2-closures of finite permutation groups, *J. London Math. Soc.* **37**, 241–252.

iebler, R.A.

[1991] The classification of distance-transitive graphs of type $q\sigma k_{q,q}$, *European J. Combin.* **12**, 125–128.

inial, N., and U. Vazirani

[1989] Graph products and chromatic numbers, in: *Proc. 30th IEEE Symp. on Foundations of Computer Science*, pp. 124–128.

ittle, C.H.C., D.D. Grant and D.A. Holton

[1975] On defect-*d* matching in graphs, *Discrete Math.* **13**, 41–54. Erratum: 1976, **14**, 203.

ivingstone, D., and A. Wagner

[1965] Transitivity of finite permutation groups on unordered sets, *Math. Z.* **90**, 393–403.

ovász, L.

[1971] On the cancellation law among finite relational structures, *Period. Math. Hungar.* **1**, 145–156.

[1972a] Direct product in locally finite categories, *Acta Sci. Math. (Szeged)* **23**, 319–322.
[1972b] A note on the line reconstruction problem, *J. Combin. Theory B* **13**, 309–310.
[1979a] *Combinatorial Problems and Exercises* (Akadémiai Kiadó/North-Holland, Budapest/Amsterdam). 2nd Edition: 1993.
[1979b] On the Shannon capacity of a graph, *IEEE Trans. Inform. Theory* **IT-25**, 1–7.

Lovász, L., and M.D. Plummer
[1986] eds., *Matching Theory, Ann. Discrete Math.* **29**.

Lubiw, A.
[1981] Some NP-complete problems similar to graph isomorphism, *SIAM J. Comput.* **10**, 11–21.

Lubotzky, A., R. Phillips and P. Sarnak
[1988] Ramanujan graphs, *Combinatorica* **8**(3), 261–277.

Luks, E.M.
[1982] Isomorphism of graphs of bounded valence can be tested in polynomial time, *J. Comput. Sys. Sci.* **25**, 42–65.
[1987] Computing the composition factors of a permutation group in polynomial time, *Combinatorica* **7**, 87–99.

Lyndon, R.C., and P.E. Schupp
[1977] *Combinatorial Group Theory* (Springer, Berlin).

Macbeath, A.M.
[1967] The classification of non euclidean plane crystallographic groups, *Canad. J. Math.* **19**, 1192–1205.

Macpherson, H.D.
[1982] Infinite distance-transitive graphs of finite valency, *Combinatorica* **2**, 63–70.

Mader, W.
[1971a] Minimale n-fach kantenzusammenhängende Graphen, *Math. Ann.* **191**, 21–28.
[1971b] Über den Zusammenhang symmetrischer Graphen, *Arch. Math.* **22**, 333–336.

Magnus, W., A. Karrass and D. Solitar
[1966] *Combinatorial Group Theory* (Interscience, New York).

Mani, P.
[1971] Automorphismen von polyedrischen Graphen, *Math. Ann.* **192**, 297–303.

Margulis, G.A.
[1988] Explicit group theoretic construction of combinatorial schemes and their application to the construction of expanders and concentrators (in Russian), *Probl. Inform. Transmission* **24**, 51–60. English translation: Plenum Press, New York, pp. 39–46.4946.

Markvorsen, S., S.Mc. Guinness and C. Thomassen
[1994] Transient random walks on graphs and metric spaces with applications to hyperbolic surfaces, to appear.

Marušič, D.
[1985] Vertex transitive graphs and digraphs of order p^k, in: *Cycles in Graphs, Burnaby, B.C., 1982*, eds. B.R. Alspach and C.D. Godsil, *North-Holland Mathematics Studies,* Vol. 115 (North-Holland, Amsterdam) pp. 115–128. MR 87b:5067.
[1987] Hamiltonian cycles in vertex symmetric graphs of order $2p^2$, *Discrete Math.* **66**, 169–174.

Marušič, D., and T.D. Parsons
[1983] Hamiltonian paths in vertex-symmetric graphs of order $4p$, *Discrete Math.* **43**, 91–96.

Maschke, H.
[1896] The representation of finite groups, especially of the rotation groups of the regular bodies of three- and four-dimensional space, by Cayley's color diagrams, *Amer. J. Math.* **18**, 156–194.

Mathon, R.
[1979] A note on the graph isomorphism counting problem, *Inform. Process. Letters* **8**, 131–132.

Matzat, B.H.
[1987] *Konstruktive Galoistheorie, Lecture Notes in Mathematics,* Vol. 1284 (Springer, Berlin).

McConnel, R.
[1963] Pseudo-ordered polynomials over a finite field, *Acta Arithm.* **8**, 127–151.

McKay, B.D.
[1987] *Nauty User's Guide*, version 1.2, Tech. Rep. tr-cs-87-03 (Department of Computer Science, Australian National University, Canberra).

McKay, B.D., and N.C. Wormald
[1984] Automorphism of random graphs with specified degrees, *Ars Combin. A* **19**, 15–26.

McKenzie, P.
[1984] Permutations of bounded degree generate groups of polynomial diameter, *Inform. Proc. Lett.* **19**, 253–254.

McKenzie, R.
[1971] Cardinal multiplication of structures with a reflexive relation, *Fund. Math.* **75**, 60–101.

Mendelsohn, E.
[1978a] Every (finite) group is the group of automorphisms of a (finite) strongly regular graph, *Ars Combin.* **6**, 75–86. MR 81a:05065.
[1978b] On the groups of automorphisms of Steiner triple and quadruple systems, *J. Combin. Theory A* **25**, 97–104. MR 80d:05010.

Mihailova, K.A.
[1958] The occurrence problem for direct products of groups, *Dokl. Akad. Nauk SSSR* **119**, 1103–1105 [1966, *Mat. Sb. (N.S.)* **70**, 241–251].

Miller, G.L.
[1979] Graph isomorphism, general remarks, *J. Comput. Syst. Sci.* **18**, 128–142.
[1983a] Isomorphism of graphs which are pairwise k-separable, *Inform. and Control* **56**, 21–33.
[1983b] Isomorphism of k-contractible graphs. A generalization of bounded valence and bounded genus, *Inform. and Control* **56**, 1–20.

Miller, G.L., and V. Ramachandran
[1992] A new graph triconnectivity algorithm and its parallelization, *Combinatorica* **12**, 53–76.

Milnor, J.
[1968a] Growth of finitely generated solvable groups, *J. Differential Geom.* **2**, 447–449.
[1968b] A note on curvature and fundamental groups, *J. Differential Geom.* **2**, 1–7.

Mnukhin, V.B.
[1987] Reconstruction of k-orbits of a permutation group, *Math. Notes* **42**, 975–980.

Mohar, B., and W. Woess
[1989] A survey on spectra of infinite graphs, *Bull. London Math. Soc.* **21**, 209–234.

Montgomery, D., and L. Zippin
[1955] *Topological Transformation Groups* (Interscience, New York).

Moon, J.W.
[1964] Tournaments with a given automorphism group, *Canad. J. Math.* **16**, 485–489. MR 29:603.

Müller, V.
[1977] The edge reconstruction hypothesis is true for graphs with more than $n \log_2 n$ edges, *J. Combin. Theory B* **22**, 281–283.

Nash-Williams, C.St.J.A.
[1978] The reconstruction problem, in: *Selected Topics in Graph Theory* (Academic Press, London) pp. 205–236.

Neumann, P.M., and C.E. Praeger
[1992] A recognition algorithm for special linear groups, *Proc. London Math. Soc.* **65**, 555–603.

Nowitz, L.A.
[1968] On the non-existence of graphs with transitive generalized dicyclic groups, *J. Combin. Theory* **4**, 49–51.

Ore, O.
[1962] ed., *Theory of Graphs, Colloq. Publ. Amer. Math. Soc.* **38**.

Oxtoby, J.C.
[1980] *Measure and Category, Graduate Texts in Mathematics*, Vol. 2 (Springer, Berlin).

Pálfy, P.P.
[1982] A polynomial bound on the orders of primitive solvable groups, *J. Algebra* **77**, 127–137.

Pólya, G.

[1937] Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen, *Acta Math.* **68**, 145 254.

Praeger, C.E.

[1985] Imprimitive symmetric graphs, *Ars Combin. A* **19**, 149–163. MR 86k:5058.

[1990] Finite vertex transitive graphs and primitive permutation groups, in: *Codes, Designs, Groups, Proc. Marshall Hall Conf., Vermont*, eds. D. Jungnickel and S.A. Vanstone (Wiley, New York) pp. 51–65.

Praeger, C.E., J. Saxl and K. Yokoyama

[1987] Distance transitive graphs and finite simple groups, *Proc. London Math. Soc.* **55**, 1–21.

Proulx, V.K.

[1977] *Classification of the toroidal groups*, Ph.D. Thesis (Columbia University).

Pultr, A., and V. Trnková

[1980] *Combinatorial, Algebraic and Topological Representations of Groups, Semigroups and Categories* (Academia Praha, Prague).

Pyber, L.

[1990] The edge reconstruction of Hamiltonian graphs, *J. Graph Theory* **14**, 173–179.

Quisquater, J.-J.

[1986] *Structures d'interconnexion: constructions et applications*, Ph.D. Thesis.

Raghunathan, M.S.

[1972] *Discrete Subgroups of Lie Groups* (Springer, New York).

Rankin, R.A.

[1948] A campanological problem in group theory, *Proc. Cambridge Philos. Soc.* **44**.

Rapaport-Strasser, E.

[1959] Cayley color groups and Hamilton lines, *Scripta Math.* **24**, 51–58.

Ringel, G., and J.W.T. Youngs

[1968] Solution of the Heawood map-coloring problem, *Proc. Nat. Acad. Sci. U.S.A.* **60**, 438–445.

Robinson, R.W.

[1970] Enumeration of non-separable graphs, *J. Combin. Theory* **9**, 327–356.

Sabidussi, G.

[1957] Graphs with given automorphism group and given graph theoretical properties, *Canad. J. Math.* **9**, 515–525.

[1960] Graph multiplication, *Math. Z.* **72**, 446 457.

[1964] Vertex-transitive graphs, *Monatsh. Math.* **68**, 426 438.

Schmerl, J.

[1979] Countable homogenous partially ordered sets, *Algebra Universalis* **9**, 317 321.

Schonland, D.S.

[1965] *Molecular Symmetry* (Van Nostrand, London).

Schwenk, A.J.

[1973] Almost all trees are cospectral, in: *New Directions in the Theory of Graphs*, ed. F. Harary (Academic Press, New York) pp. 275–308.

Serre, J.-P.

[1977] *Arbres, Amalgames, SL_2, Astérisque*, Vol. 46 (Société Mathématique de France, Paris).

[1980] *Trees* (Springer, Berlin).

Shelah, S.

[1980] On a problem of Kurosh, Jónsson groups and applications, in: *Word Problems II*, eds. S.I. Adian, W.W. Boone and G. Higman (North-Holland, Amsterdam) pp. 373–394.

Sims, C.C.

[1967] Graphs and finite permutation groups, *Math. Z.* **95**, 76–86.

[1970] Computational methods in the study of permutation groups, in: *Computational Problems in Abstract Algebra*, ed. J. Leech (Pergamon Press, Oxford) pp. 169–183.

[1971] Computation with permutation groups, in: *Proc. 2nd Symp. on Symbolic Algebraic Manipulation* (ACM, New York) pp. 23–28.

[1978] *Some Group Theoretic Algorithms, Lecture Notes in Mathematics*, Vol. 697 (Springer, Berlin) pp. 108-124.

Smith, D.H.

[1974] Distance-transitive graphs of valency four, *J. London Math. Soc.* **8**, 377–384.

Spencer, J.

[1983] What's not inside a Cayley graph? *Combinatorica* **3**, 239–241.

Stallings, J.

[1971] *Group Theory and Three-Dimensional Manifolds* (Yale University Press, New Haven, CT).

Stockmeyer, P.K.

[1977] The falsity of the reconstruction conjecture for tournaments, *J. Graph Theory* **1**, 19–25.

Tardos, G.

[1992] Intersection of subgroups of a free group, *Invent. Math.* **108**, 29–36.

Thomas, S.

[1987] The infinite case of the edge-orbit conjecture, *Algebra Universalis* **24**, 167–168.

Thomassen, C.

[1983] Girth in graphs, *J. Combin. Theory B* **35**, 129–141.

[1990] Resistances and currents in infinite electrical networks, *J. Combin. B* **49**, 87–102.

[1991] Tilings of the torus and the Klein bottle and vertex-transitive graphs on a fixed surface, *Trans. Amer. Math. Soc.* **323**, 605–635.

[1992] The Hadwiger number of infinite vertex-transitive graphs, *Combinatorica* **12**, 481–491.

Thomassen, C., and W. Woess

[1994] Vertex-transitive graphs and accessibility, *J. Combin. Theory B* **26**, 1–60.

Thompson, J.G.

[1984] Some finite groups which appear as $Gal(L/K)$ where $K \subseteq Q(\mu_n)$, *J. Algebra* **89**, 437–499.

Titov, V.K.

[1975] On the symmetry of graphs, in: *Proc. 2nd All-Union Seminar on Combinatorial Mathematics, Moscow, Voprosy Kibernetiki* 15, 76–109.

Tits, J.

[1959] Sur la trialité at certains groupes qui s'en déduisent, *Publ. Math. IHES* **2**, 14–60.

[1970] Sur le groupe des automorphismes d'un arbre, in: *Essays on Topology and Related Topics, Mémoires dédiés a Georges de Rham* (Springer, Berlin) pp. 188–211.

[1972] Free subgroups of linear groups, *J. Algebra* **20**, 250–270.

Trofimov, V.I.

[1985] Graphs with polynomial growth, *Math. USSR Sbornik* **51**, 405–417.

[1992] On the action of a group on a graph, *Acta Appl. Math.* **29**, 161–170.

Truss, J.K.

[1992] Generic automorphisms of homogeneous structures, *Proc. London Math. Soc.* **65**, 121–141.

Tucker, T.W.

[1983] Finite groups acting on surfaces and the genus of a group, *J. Combin. Theory B* **34**, 82–98.

[1984] On Proulx's four exceptional toroidal groups, *J. Graph Theory* **8**, 29–33.

Tutte, W.T.

[1947] A family of cubical graphs, *Proc. Cambridge Philos. Soc.* **43**, 459–474.

[1979] All the king's horses, in: *Graph Theory and Related Topics*, eds. J.A. Bondy and U.S.R. Murty (Academic Press, New York) pp. 15–33.

van Bon, J., and A.E. Brouwer

[1987] The distance-regular antipodal covers of classical distance-regular graphs, *Coll. Math. Soc. János Bolyai* **52**, 141–166.

Varopoulos, N.Th.

[1985] Isoperimetric inequalities and Markov chains, *J. Funct. Anal.* **63**, 215–239.

[1991] Analysis and geometry on groups, in: *Proc. Int. Congr. of Mathematicians, Kyoto, 1990* (Springer, Kyoto) pp. 951–957.

Vopenka, P., A. Pultr and Z. Hedrlín

[1965] A rigid relation exists on any set, *Comment. Math. Univ. Carol.* **6**, 149–155.

Watkins, M.E.

[1970] Connectivity of transitive graphs, *J. Combin. Theory* **8**, 23–29.

[1971] On the action of non-abelian groups on graphs, *J. Combin. Theory B* **1**, 95–104.

Weisfeiler, R.

[1976] *On Construction and Identification of Graphs, Lecture Notes in Mathematics*, Vol. 556 (Springer, Berlin).

Weiss, R.M.

[1976] Über lokal *s*-reguläre Graphen, *J. Combin. Theory B* **20**, 124–127.

[1979] Elations of graphs, *Acta Math. Acad. Sci. Hungar.* **34**, 101–103.

[1981] The nonexistence of 8-transitive graphs, *Combinatorica* **1**, 309–311. MR 84f:05050.

[1985a] Distance-transitive graphs and generalized polygons, *Arch. Math. (Basel)* **45**, 186–192. MR 87a:05081.

[1985b] On distance transitive graphs, *Bull. London Math. Soc.* **17**, 253–256.

White, A.T.

[1973] *Graphs, Groups, and Surfaces* (North-Holland, Amsterdam).

[1985] Ringing the changes II, *Ars Combin. A* **20**, 65–75. MR 87c:05080.

Whitney, H.

[1932] Congruent graphs and the connectivity of graphs, *Amer. J. Math.* **54**, 150–168 [see Lovász 1979a, Problem 15.1].

Wielandt, H.

[1964] *Finite Permutation Groups* (Academic Press, New York).

[1969] *Permutation Groups through Invariant Relations, Lecture Notes* (Ohio State University).

Wigner, E.

[1927] Einige Folgerungen aus der Schrödingerschen Theorie für die Termstrukturen, *Z. Phys.* **43**, 624–652.

[1930] Über die elastischen Eigenschwingungen symmetrischer Systeme, *Nachr. Ges. Wiss. Göttingen, Math.-Phys. Klasse* **133**.

Wigner, E.P.

[1959] *Group Theory and its Application to the Quantum Mechanics of Atomic Spectra* (Academic Press, New York).

Wilkie, H.C.

[1966] On non-euclidean crystallographic groups, *Math. Z.* **91**, 87–102.

Wilson, R.M.

[1974] Nonisomorphic Steiner triple systems, *Math. Z.* **135**, 303–313.

Witte, D.

[1986] Cayley digraphs of prime-power order are Hamiltonian, *J. Combin. Theory B* **40**, 107–112.

Witte, D., and J.A. Gallian

[1984] A survey: Hamiltonian cycles in Cayley graphs, *Discrete Math.* **51**, 293–304.

Wolf, J.A.

[1968] Growth of finitely generated solvable groups and curvature of Riemannian manifolds, *J. Differential Geometry* **2**, 421–446.

Wolf, T.R.

[1982] Solvable and nilpotent subgroups of GL(n,qm), *Canad. J. Math.* **34**, 1097–1111.

Wong, W.J.

[1967] Determination of a class of permutation groups, *Math. Z.* **99**, 235–246.

Woodrow, R.E.

[1979] There are four countable ultrahomogeneous graphs without triangles, *J. Combin. Theory B* **27**, 168–179.

Wormald, N.

[1986] A simpler proof of the asymptotic formula for the number of unlabelled *r*-regular graphs, *Indian J. Math.* **28**, 43–47.

Wright, E.M.
[1971] Graphs on unlabelled nodes with a given number of edges, *Acta Math.* **126**, 1–9.

Zemlyachenko, V.M., N.M. Kornienko and R.I. Tyshkevich
[1985] Graph isomorphism problem, *J. Soviet Math.* **29**, 1426–1481. Original: 1982, *Zapiski LOMI* **118**.

Zieschang, H., E. Vogt and H.-D. Coldewey
[1980] *Surfaces and Planar Discontinuous Groups, Lecture Notes in Mathematics*, Vol. 835 (Springer, Berlin).

CHAPTER 28

Combinatorial Optimization

M. GRÖTSCHEL

Konrad-Zuse-Zentrum für Informationstechnik, Heilbronner Str. 10, D-10711 Berlin-Wilmersdorf, Germany

L. LOVÁSZ

Department of Computer Science, Yale University, New Haven, CT 06250, USA

Contents

HANDBOOK OF COMBINATORICS
Edited by R. Graham, M. Grötschel and L. Lovász

1. Introduction

Optimizing means finding the maximum or minimum of a certain function, called the objective function, defined on some domain. Classical theories of optimization (differential calculus, variational calculus, optimal control theory) deal with the case when this domain is infinite. From this angle, the subject of combinatorial optimization, where the domain is typically finite, might seem trivial: it is easy to say that "we choose the best from this finite number of possibilities". But, of course, these possibilities may include all trees on *n* nodes, or all Hamilton circuits of a complete graph, and then listing all possibilities to find the best among them is practically hopeless even for instances of very moderate size. In the framework of complexity theory (chapter 29), we want to find the optimum in polynomial time. For this (or indeed, to do better than listing all solutions) the special structure of the domain must be exploited.

Often, when the objective function is too wild, the constraints too complicated, or the problem size too large, it is impossible to find an optimum solution. This is quite frequently not just a practical experience; mathematics and computer science have developed theories to make intuitive assertions about the difficulty of certain problems precise. Foremost of these is the theory of NP-completeness (see chapter 29).

In cases when optimum solutions are too hard to find, algorithms (so-called *heuristics*) can often be designed that produce approximately optimal solutions. It is important that these suboptimal solutions have a guaranteed quality; e.g., for a given maximization problem, the value of the heuristic solution is at least 90% of the optimum value for every input.

While not so apparent on the surface, of equal importance are algorithms, called *dual heuristics*, that provide (say for a maximization problem again) upper bounds on the optimum value. Dual heuristics typically solve so-called *relaxations* of the original problem, i.e., optimization problems that are obtained by dropping or relaxing the constraints so as to make the problem easier to solve. Bounds computed this way are important in the analysis of heuristics, since (both theoretically and in practice) one compares the value obtained by a heuristic with the value obtained by the dual heuristic (instead of the true optimum, which is unknown). Relaxations also play an important role in general solution schemes for hard problems like branch-and-bound.

The historical roots of combinatorial optimization lie in problems in economics: the planning and management of operations and the efficient use of resources. Soon more technical applications came into focus and were modelled as combinatorial optimization problems, such as sequencing of machines, scheduling of production, design and layout of production facilities. Today we see that discrete optimization problems abound everywhere. We encounter them in areas such as portfolio selection, capital budgeting, design of marketing campaigns, investment planning and facility location, political districting, gene sequencing, classification of plants and animals, the design of new molecules, the determination of ground states, layout of survivable and cost-coefficient communication networks, positioning of satellites, design and production of VLSI circuits and printed circuit boards, sizing of truck fleets and transportation planning, layout of mass transportation systems and scheduling of buses and trains, assignment of workers to jobs such as drivers to buses and airline crew scheduling, design of unbreakable codes, etc. The list of applications seems endless; even in areas like sports, archeology or psychology, combinatorial optimization is used to answer important questions. We refer to chapter 35 for a detailed description of several real-world examples.

There are basically two ways of presenting "combinatorial optimization": by problems or by methods. Since combinatorial optimization problems abound in this Handbook and many chapters deal with particular problems, discuss their practical applications and algorithmic solvability, we organize our material according to the second approach. We will describe the fundamental algorithmic techniques in detail and illustrate them on problems to which these methods have been applied successfully.

Some important aspects of combinatorial optimization algorithms we can only touch in this chapter. One such aspect is *parallelism*. There is no doubt that the computers of the future will be parallel machines. A systematic treatment of parallel algorithms is difficult since there are many computer architectures, based on different principles, and each architecture leads to a different model of parallel computational complexity. One very general model of parallel computation is described in chapter 29.

Another important aspect is the *on-line* solution of combinatorial problems. We treat here "static" problems, where all data are known before the optimization algorithm is called. In many practical situations, however, data come in one by one, and decisions must be made before the next piece of data arrives. The theoretical modelling of such situations is difficult, and we refrain from discussing the many possibilities.

A third disclaimer of this type is that we focus on the *worst-case* analysis of algorithms. From a practical point of view, *average-case analysis*, i.e., the analysis of algorithms on random input, would be more important; but for a mathematical treatment of this, one has to make assumptions about the input distribution, and except for very simple cases, such assumptions are extremely difficult to justify and lead to unresolvable controversies.

Finally, we should mention the increasing significance of *randomization*. It is pointed out in chapter 29 that randomization should not be confused with the average case analysis of algorithms. Randomization is often used in conjunction with determinant methods, and it is an important tool in avoiding "traps" (degeneracies), in reaching tricky "corners" of the domain, and in many other situations. We do discuss several of these methods; other issues like *derandomization* (transforming randomized algorithms into deterministic ones), or the reliability of random number generators will, however, not be treated here.

Combinatorial optimization problems typically have as inputs methods based on determinant computation, combinatorial structures and numbers (e.g., a graph with weights on the edges). In the Turing machine model, both are encoded as

0–1 strings; but in the RAM machine model it is natural to consider the input as a sequence of integers. If the input also involves a combinatorial structure, then the combinatorial structure can be considered as a set of 0s and 1s. We denote by $\langle a \rangle$ the number of bits in the binary representation of the integer a; for a matrix $A = (a_{ij})$ of integers, we define $\langle A \rangle = \sum \langle a_{ij} \rangle$.

An algorithm (with an input sequence of integers $a_1, \ldots, a_n\rangle$ runs in *polynomial time* (short: is *polynomial*) if it can be implemented on the RAM machine so that the number of bit-operations performed is bounded by a polynomial in the number of input bits $\langle a_1 \rangle + \cdots + \langle a_n \rangle$. Considering the input as a set of numbers allows two important versions of this notion.

An algorithm (with an input sequence of integers $a_1, \ldots, a_n$) is *pseudopolynomial*, if it can be implemented on the RAM machine so that the number of bit-operations performed is bounded by a polynomial in $|a_1| + \cdots + |a_n|$. (This can also be defined on the Turing machine model: it corresponds to polynomial running time where the encoding of an integer a by a string of a 1s is used. Thus a pseudopolynomial algorithm is also called *polynomial in the unary encoding*.) Clearly, every polynomial algorithm is pseudopolynomial, but not the other way around: testing primality in the trivial way by searching through all smaller integers is pseudopolynomial but not polynomial.

An algorithm (with input integers $a_1, \ldots, a_n$) is *strongly polynomial* if it can be implemented on the RAM machine in $O(n^c)$ steps with numbers of $O((\langle a_1 \rangle + \cdots + \langle a_n \rangle)^c)$ digits for some $c > 0$. Clearly every strongly polynomial algorithm is polynomial, but not the other way around: e.g., the Euclidean algorithm is polynomial but not strongly polynomial. On the other hand, Kruskal's algorithm for shortest spanning trees is strongly polynomial.

Further reading. Bachem et al. (1983), Ford and Fulkerson (1962), Gondran and Minoux (1979), Grötschel et al. (1988), Lawler et al. (1985), Nemhauser et al. (1989), Nemhauser and Wolsey (1988), Schrijver (1986).

2. The greedy algorithm

Kruskal's algorithm and matroids. The most natural principle we can try to build an optimization algorithm on is *greediness*: building up the solution by making the best choice locally. As the most important example of an optimization problem where this simple idea works, let us recall Kruskal's algorithm for a shortest spanning tree from chapters 2, 9 and 40.

Given a connected graph G with n nodes, and a length function $c: E(G) \rightarrow \mathbb{Z}$, we want to find a shortest spanning tree (where the length of a subgraph of G is defined as the sum of the lengths of its edges). The algorithm constructs the tree by finding its edges $e_1, e_2, \ldots, e_{n-1}$ one by one:

- e_1 is an edge of minimum length;
- e_k is an edge such that $e_k \not\in \{e_1, \ldots, e_{k-1}\}$, $\{e_1, \ldots, e_k\}$ is a forest and

$$c(e_k) = \min\{c(e) \mid e \not\in \{e_1, \ldots, e_{k-1}\} \text{ and } \{e_1, \ldots, e_{k-1}, e\} \text{ is a forest}\} .$$

Each step of the algorithm makes locally the best choice: this is why it is called a *greedy* algorithm. Kruskal's theorem (going back actually to Borůvka in 1926; see Graham and Hell (1985) for an account of its history) asserts that $\{e_i: 1 \leq i \leq n-1\}$ is a shortest spanning tree of G, i.e., *the spanning tree constructed by the greedy algorithm is optimal.*

The graph structure plays little role in the algorithm; the only information about the graph used is that "$\{e_1, \ldots, e_{k-1}, e\}$ is a forest". In fact, this observation is one of the possible routes to the notion of *matroids.*

Let S be a finite set, $c: S \to \mathbb{Z}_+$, a cost function on S, and $\mathcal{F} \subseteq 2^S$ a *hereditary family* (independence system) of subsets of S, i.e., a set of subsets such that $X \in \mathcal{F}$ and $Y \subseteq X$ implies $Y \in \mathcal{F}$. The goal is to find $\max\{c(X) = \sum_{e \in X} c(e) \mid X \in \mathcal{F}\}$. In this more general setting, the greedy algorithm can still be easily formulated. It constructs a maximal set X by finding its elements $e_1, e_2, \ldots$ one by one, as follows:

- e_1 is an element of maximum cost in $\bigcup_{X \in \mathcal{F}} X$,
- e_k is defined such that $e_k \not\in \{e_1, \ldots, e_{k-1}\}$, $\{e_1, \ldots, e_k\} \in \mathcal{F}$ and

$$c(e_k) = \max\{c(e) \mid e \not\in \{e_1, \ldots, e_{k-1}\}, \text{ and } \{e_1, \ldots, e_{k-1}, e\} \in \mathcal{F}\}\ .$$

The algorithm terminates when no such element exists. We call a set $X_{\mathrm{gr}} := \{e_1, \ldots, e_r\}$ obtained by this algorithm a *greedy solution*, and let X_{opt} denote an optimum solution to our problem. In chapter 9 it is shown that the greedy solution is optimal for every cost function if and only if $(E, \mathcal{F})$ is a matroid.

There are other problems where the greedy algorithm (with an appropriate interpretation) gives an optimum solution. The notion of *greedoids* (see chapter 9, or Korte et al. 1991) is an attempt to describe a general class of such problems. Further examples are polymatroids (see chapter 11), coloring of various classes of perfect graphs (see chapter 4) etc. In fact, an optimization model where the greedy solution is optimal was described by Monge in 1781!

Greedy heuristic. But in most optimization problems, greed does not pay: the greedy solution is not optimal in general. We may still use, however, a greedy algorithm as a heuristic, to obtain a "reasonable" solution (and in practice it is very often used indeed).

We measure the quality of the heuristic by comparing the value of the objective function at the heuristic solution with its value at the optimum solution. To be precise, consider, say, a minimization problem. For convenience, let us assume that every instance has a positive optimum objective function value v_{opt}, say. For a given instance, let v_{heur} denote the objective value achieved by a given heuristic (if the heuristic includes free choices at certain points, we define v_{heur} as the value achieved by the worst possible choices). We define the *performance ratio* of a heuristic as the supremum of $v_{\mathrm{heur}}/v_{\mathrm{opt}}$ over all problem instances. The *asymptotic performance ratio* is the lim sup of this ratio, assuming that $v_{\mathrm{opt}} \to \infty$. For a maximization problem, we replace this quotient by $v_{\mathrm{opt}}/v_{\mathrm{heur}}$.

Rank quotients. Consider a hereditary family $\mathcal{F} \subseteq 2^E$ and a weight function $c: E \to \mathbb{Z}_+$ on E again. We want to find a maximum weight member of $\mathcal{F}$. Let the greedy algorithm, just as above, give a set X_{gr}, and let X_{opt} denote an optimum solution.

Since we are maximizing, trivially

$$c(X_{\text{gr}}) \leq c(X_{\text{opt}}) .$$

Define, for $X \subseteq E$, the *upper rank of* X by

$$r(X) = \max\{|Y|: Y \subseteq X, Y \in \mathcal{F}\} .$$

We also define the *lower rank* of X by

$$\rho(X) = \min\{|Y|: Y \subseteq X, Y \in \mathcal{F}, \nexists U \in \mathcal{F} \text{ with } Y \subset U \subseteq X\} .$$

Note that matroids are just those hereditary families with $r = \rho$. In general, define the *rank quotient* of $(E, \mathcal{F})$ by

$$\gamma_{\mathcal{F}} = \min\left\{\frac{\rho(X)}{r(X)}: X \subseteq E, r(X) > 0\right\} .$$

Note that $\gamma_{\mathcal{F}}$ is just the worst ratio between $c(X_{\text{gr}})$ and $c(X_{\text{opt}})$ for 0–1 valued weight functions c. The following theorem of Jenkyns (1976) and Korte and Hausmann (1978) gives a performance guarantee for the greedy algorithm by showing that 0–1 weightings are the worst case.

Theorem 2.1. *For every hereditary family* $(E, \mathcal{F})$ *and every weight function* $c: E \to \mathbb{Z}_+$, *we have*

$$c(X_{\text{gr}}) \geq \gamma_{\mathcal{F}} c(X_{\text{opt}}) .$$

As an application, consider the greedy heuristic for the matching problem. Here G is a graph, $E = E(G)$, and $\mathcal{F}$ consists of all matchings, i.e., sets of edges with no common endnode. We claim that $\gamma_{\mathcal{F}} \geq \frac{1}{2}$. In fact, if $X \subseteq E$ and M is a smallest non-extendible matching in X then the $2|M|$ endpoints of the edges in M cover all edges in X, and hence a maximum matching in X cannot have more than $2|M|$ edges. Thus we obtain that *for every weighting, the greedy algorithm for the matching problem has performance ratio at most* 2.

Greedy blocking sets. We discuss a greedy heuristic with a somewhat more involved analysis. Let $(V, \mathcal{H})$ be a hypergraph. A *blocking set* or *cover* of the hypergraph is a subset $S \subseteq V$ that meets (blocks) every member of $\mathcal{H}$. Let S_{opt} denote a blocking set with a minimum number of elements, and define the *covering (or blocking) number* by $\tau(\mathcal{H}) := |S_{\text{opt}}|$. (To compute $\tau(\mathcal{H})$ is NP-hard in general; cf. also chapters 7 and 24.)

A *greedy blocking set* S_{gr} is constructed as follows. Choose a vertex v_1 of maximum degree. If $v_1, \ldots, v_k$ have been selected, choose v_{k+1} to block as many members of $\mathcal{H}$ not blocked by $v_1, \ldots, v_k$ as possible. We stop when all members of $\mathcal{H}$ are blocked.

Concerning the performance of this algorithm, we have the following bound (Johnson 1974, Stein 1974, Lovász 1975).

Theorem 2.2. *Let $\mathcal{H}$ be a hypergraph with maximum degree Δ. Then for the size of any greedy blocking set S_{gr},*

$$|S_{\mathrm{gr}}| \leq \left(1 + \frac{1}{2} + \cdots + \frac{1}{\Delta}\right)\tau(\mathcal{H}) .$$

(The "error factor" on the right-hand side is less than $1 + \ln \Delta$. It is easy to see that the ratio $1 + 1/2 + \cdots + 1/\Delta$ cannot be improved.)

Proof. Let k_i denote the number of vertices in S_{gr} selected in the phase when we were able to block exactly i new edges at a time. Let $\mathcal{H}_i$ denote the set of edges first blocked in this phase. So $|\mathcal{H}_i| = ik_i$, and

$$|S_{\mathrm{gr}}| = k_\Delta + k_{\Delta-1} + \cdots + k_1 .$$

Now consider the optimum blocking set S_{opt}. Every vertex in S_{opt} (in fact, every vertex of $\mathcal{H}$), blocks at most i edges from $\mathcal{H}_i \cup \mathcal{H}_{i-1} \cup \cdots \cup \mathcal{H}_1$, since a vertex blocking more from this set should have been included in the greedy blocking set before phase i. Hence

$$|S_{\mathrm{opt}}| \geq \frac{1}{i}\,(ik_i + (i-1)k_{i-1} + \cdots + k_1) . \tag{1}$$

Multiplying this inequality by $1/(i+1)$ for $i = 1, \ldots, \Delta - 1$, and by 1 for $i = \Delta$, and then adding up the resulting inequalities, we obtain that

$$\left(1 + \frac{1}{2} + \cdots + \frac{1}{\Delta}\right)|S_{\mathrm{opt}}| \geq k_1 + \cdots + k_\Delta = |S_{\mathrm{gr}}| . \qquad \square$$

Note that in this proof, we do not directly compare $|S_{\mathrm{gr}}|$ with $|S_{\mathrm{opt}}|$; the latter is not available. Rather, we use a family (1) of lower bounds on $|S_{\mathrm{opt}}|$. For each i, $(1/i)(\mathcal{H}_1 \cup \cdots \cup \mathcal{H}_i)$ can be viewed as a *fractional matching* (see chapters 7, 24), and so the theorem can be sharpened by using the fractional matching (or cover) number τ^* on the right-hand side. In fact, the fractional cover number τ^* is a relaxation of the cover number τ, obtained by formulating τ as the optimum value of an integer linear program and dropping the integrality constraints.

Greedy travelling salesman tours. Recall the travelling salesman problem (TSP): given a graph $G = (V, E)$ on $n \geq 3$ nodes, and "distances" $c: E \to \mathbb{Z}_+$, find a Hamilton circuit with minimum length. Usually it is not an essential restriction of generality to assume that G is a complete graph. Moreover, in many important applications, the distance function c satisfies the triangle inequality: $c_{ij} + c_{jk} \geq c_{ik}$ for any three distinct nodes i, j, and k. We shall always restrict ourselves to the special case of the complete graph with lengths satisfying the triangle inequality (called the *metric case*).

The travelling salesman problem is NP-hard. Linear programming provides a practically quite efficient method to solve it (cf. section 8). Here we discuss two greedy heuristics.

The first one, called NEAREST NEIGHBOR heuristic, is an obvious idea: Choose some arbitrary node and visit it; from the last node visited go to the closest not yet visited node; if all nodes have been visited, return to the first node. This heuristic does make locally good choices, but it may run into traps. Figure 2.1 shows the result of a NEAREST NEIGHBOR run for a TSP consisting of 52 points of interest in Berlin, starting at point 1, the Konrad-Zuse-Zentrum. It is clear that this is far from being optimal. In fact, series of metric n-city TSP instances can be constructed where the tour built by the NEAREST NEIGHBOR heuristic is $\Omega(\log n)$ as long as the optimum tour length (Rosenkrantz et al. 1977).

The second greedy heuristic, called NEAREST INSERTION, will never show such a poor performance. It works as follows. We build up a circuit going through more and more nodes.

- Start with any node v_1 (viewed as a circuit with one node).
- Let T_k be a circuit of length k already constructed. Choose a node v_k not on T_k and a node u_k on T_k such that the distance $c_{u_k v_k}$ is minimal. Delete one of the edges of T_k incident with u_k, and connect its endpoints to v_k, to get a circuit T_{k+1}.
- After n steps we get a Hamiltonian circuit T_{nins}.

Another way to describe this heuristic is the following. The tree F formed by

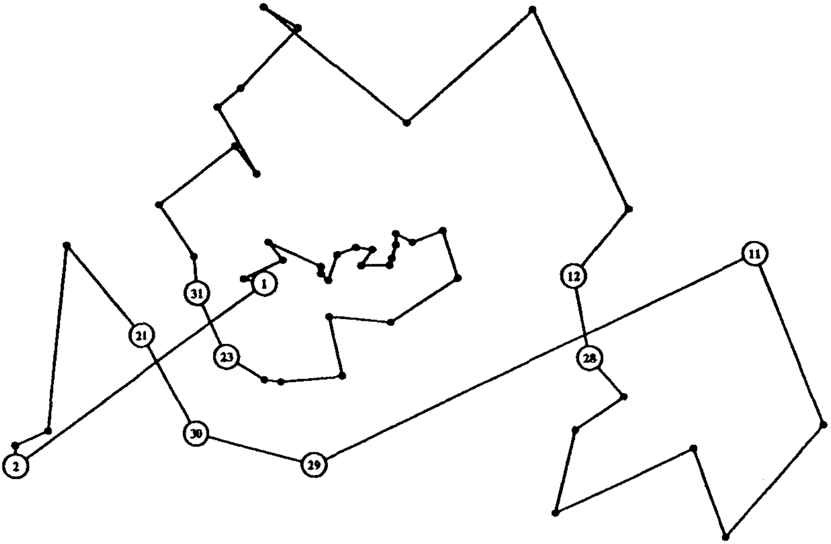


Figure 2.1. A nearest neighbor tour of a 52-city-problem.

the edges $v_k u_k$ is constructed by Prim's algorithm, (see, e.g., Cormen et al. 1990) and so it is a shortest spanning tree of G. We double each edge of F to obtain an Eulerian graph. An Euler tour of this visits every vertex at least once; making shortcuts, we obtain a tour that visits all nodes exactly once.

Theorem 2.3. *T_{nins} is at most twice as long as the optimum tour.*

Proof. Let T_{opt} denote an optimum tour. We make two observations. First, deleting any edge from T_{opt} we get a spanning tree. Hence

$$c(T_{\text{opt}}) \geq c(F)\,. \tag{2}$$

Second, inserting v_k into T_k we increase T_k's length by at most $2c_{u_k v_k}$; adding up these increments, we get

$$c(T_{\text{nins}}) \leq 2c(F)\,. \qquad \square$$

Note that this argument (implicitly) uses a *dual heuristic* also: inequality (2) makes use of the fact that the length of the shortest spanning tree is a lower bound on the length of the shortest tour. The value of this dual heuristic is easily found by the greedy algorithm. In fact, a somewhat better lower bound could be obtained by considering *unicyclic* subgraphs, i.e., those subgraphs containing at most one circuit. The edge-sets of such subgraphs also form a matroid. The bases of this matroid are connected spanning subgraphs containing exactly one circuit; for this chapter, we call such subgraphs 1-trees. A 1-tree with minimum length can be found easily by the greedy algorithm, and since every tour is a 1-tree, this gives a lower bound on the minimum tour. We will see in section 9 how to further improve this lower bound.

There are several other greedy-like heuristics for the travelling salesman problem, such as farthest insertion, cheapest insertion, sweep, savings etc. Many of them do not have, however, a proven constant performance ratio, although some of them work better in practice than the nearest insertion heuristic (see Reinelt 1994).

The heuristic for the traveling salesman problem with the best known performance ratio is due to Christofides (1976). It uses the shortest spanning tree F, but instead of doubling its edges to make it Eulerian, it adds a matching M with minimum length on the set of nodes with odd degree in F. (Such a matching can be found in polynomial time, see chapter 3.) It is easy to see that $2c(M) \leq c(T_{\text{opt}})$, and hence the Christofides heuristic has performance ratio of at most $\frac{3}{2}$.

Bin packing. Let $a_1, \ldots, a_n \leq 1$ be positive real numbers ("weights"). We would like to partition them into classes ("bins") $B_1, \ldots, B_k$ so that the total weight of every bin is at most 1. Our aim is to minimize the number k of bins. Let k_{opt} be this minimum. To compute k_{opt} is NP-hard.

A trivial lower bound on how well we can do is the roundup of the total weight $w := \sum_i a_i$. The following simple (greedy) heuristic already gets asymptotically within a factor of 2 to this lower bound.

NEXT-FIT HEURISTIC. We process $a_1, a_2, \ldots$ one by one. We put them into one bin until putting the next a_i into the bin would increase its weight over 1. Then we close the bin and start a new one. We denote by k_{nf} the number of bins used by this heuristic.

Theorem 2.4. $k_{nf} < 2w + 1$; $k_{nf} < 2k_{\text{opt}}$.

Proof. Let $k := k_{\text{nf}}$, and denote by w_i the weight put in bin B_i by the NEXT-FIT heuristic ($1 \leq i \leq k$). Then clearly $w_i + w_{i+1} > 1$ (since otherwise the first weight in B_{i+1} should have been put in B_i). If k is even, this implies

$$w = (w_1 + w_2) + (w_3 + w_4) + \cdots + (w_{k-1} + w_k) > k/2 ,$$

and hence

$$k < 2w \leq 2k_{\text{opt}} .$$

If k is odd, we obtain

$$\begin{aligned} w &= \tfrac{1}{2}(w_1 + (w_1 + w_2) + (w_2 + w_3) + \cdots + (w_{k-1} + w_k) + w_k) \\ &\geq \tfrac{1}{2}(k - 1 + w_1 + w_2) > \frac{k-1}{2} , \end{aligned}$$

whence

$$k < 2w + 1 \leq 2k_{\text{opt}} + 1 ,$$

and hence, $k \leq 2k_{\text{opt}} - 1 < 2k_{\text{opt}}$. □

The following better heuristic is still very "greedy".

FIRST-FIT HEURISTIC. We process $a_1, a_2, \ldots$ one by one, putting each a_i into the first bin into which it fits. We denote by k_{ff} the number of bins used by this heuristic.

The following bound on the performance of FIRST FIT, due to Garey et al. (1976), is substantially more difficult to prove than the previous one.

Theorem 2.5. $k_{\text{ff}} \leq \lceil \frac{17}{10} k_{\text{opt}} \rceil$. *There exist lists with arbitrarily large total weight for which* $k_{\text{ff}} \geq \lceil \frac{17}{10} k_{\text{opt}} \rceil - 1$.

A natural (still "greedy") improvement on the FIRST-FIT heuristic is to preprocess the weights by ordering them decreasingly, and then apply FIRST-FIT. We call this the FIRST-FIT DECREASING heuristic, and denote the number of bins it uses by k_{ffd}. This preprocessing does improve the performance, as shown by the following theorem of Johnson (1973).

Theorem 2.6. $k_{\text{ffd}} \leq \lceil \frac{11}{9} k_{\text{opt}} \rceil + 4$. *There exist lists with arbitrarily large total weight for which* $k_{\text{ffd}} = \lceil \frac{11}{9} k_{\text{opt}} \rceil$.

There are other greedy-like heuristics for bin packing, e.g., best fit, whose performance ratio is similar; the heuristic called harmonic fit is slightly better. (See Coffman et al. 1984 for a survey.) More involved heuristic algorithms for the bin packing problem use linear programming and achieve an asymptotic performance ratio arbitrarily close to 1; see section 8.

The first two heuristics above are special in the sense that they are *on-line*: each weight is placed in a bin without knowing those that follow, and once a weight is placed in a bin, it is never touched again. On-line heuristics cannot achieve as good a performance ratio as general heuristics; it is easy to show that for any on-line heuristic for the bin packing problem there exists a sequence of weights (with arbitrarily large total weight) for which it uses at least $\frac{3}{2} k_{\text{opt}}$ bins. This lower bound can be improved to 1.54, van Vliet (1992).

The knapsack problem. Given a knapsack with total capacity b, and n objects with weights $a_1, \ldots, a_n$ of value $c_1, \ldots, c_n$, we want to pack as much value into the knapsack as possible. In other words, given positive integers $b, a_1, \ldots, a_n, c_1, \ldots, c_n$, we want to maximize $\sum_i c_i x_i$ subject to the constraints $\sum i\, a_i x_i \leq b$ and $x_i \in \{0, 1\}$, $i = 1, \ldots, n$. To exclude trivial cases, we may assume that $a_1, \ldots, a_n \leq b$. We denote by C_{opt} the optimum value of the objective function.

The knapsack problem is NP-hard; we will see in section 9, however, that one can get arbitrarily close to the optimum in polynomial time. Here we describe a greedy algorithm that has performance ratio at most 2.

It is clear that we want to select objects with small weight and large value; so it is natural to put in the knapsack an object with largest "weight density" c_i/a_i. Assume that the objects are labelled so that $c_1/a_1 \geq c_2/a_2 \geq \cdots \geq c_n/a_n$. Then the greedy algorithm consists of selecting $x_1, x_2, \ldots, x_n \in \{0, 1\}$ recursively as follows:

$$x_i = \begin{cases} 1\,, & \text{if } a_i \leq b - \sum_{j=1}^{i-1} a_j x_j\,, \\ 0\,, & \text{otherwise}\,. \end{cases}$$

Theorem 2.7. *For the solution* $x_1, \ldots, x_n$ *obtained by the (weight density) greedy algorithm, we have*

$$\sum_{i=1}^{n} c_i x_i \geq C_{\text{opt}} - \max_i c_i\,.$$

It follows that comparing the greedy solution with the best of the trivial solutions (select one object), we get a heuristic with performance ratio at most 2.

Proof. Let x be the solution found by the greedy algorithm, and y be an optimum solution. If x is not optimal, there must be an index j such that $x_j = 0$ and $y_j = 1$;

consider the least such j. Then we have for the greedy solution,

$$\begin{aligned} C_{\mathrm{gr}} := \sum_{i=1}^{n} c_i x_i \geq \sum_{i=1}^{j} c_i x_i &= \sum_{i=1}^{j} c_i y_i + \sum_{i=1}^{j} c_i (x_i - y_i) \\ \geq \sum_{i=1}^{j} c_i y_i + \sum_{i=1}^{j} \frac{c_j}{a_j} a_i (x_i - y_i) &= \sum_{i=1}^{j} c_i y_i + \frac{c_j}{a_j} \Big(\sum_{i=1}^{j} a_i x_i - \sum_{i=1}^{j} a_i y_i \Big) . \end{aligned} \tag{3}$$

Since j was not chosen by the greedy algorithm, we have here

$$\sum_{i=1}^{j} a_i x_i \geq b - a_j .$$

Furthermore, the feasibility of y implies that

$$\sum_{i=1}^{j} a_i y_i \leq b - \sum_{i=j+1}^{n} a_i y_i .$$

Substituting in (3), we obtain

$$C_{\mathrm{gr}} \geq \sum_{i=1}^{j} c_i y_i + \frac{c_j}{a_j} \Big(\sum_{i=j+1}^{n} a_i y_i - a_j \Big) \geq \sum_{i=1}^{j} c_i y_i - c_j = C_{\mathrm{opt}} - c_j . \qquad \square$$

3. Local improvement

The greedy algorithm and its problem-specific versions belong to a class of algorithms sometimes called (one-pass) *construction heuristics*. A myopic rule is applied and former decisions are not reconsidered. The purpose of these heuristics is to find a "reasonable" feasible solution very fast. But, of course, a locally good choice may lead to a globally poor solution.

Exchange heuristics for the travelling salesman problem. We have seen such unpleasant behavior in fig. 2.1: initially, short connections are chosen but, in the end, very long steps have to be made to connect the "forgotten" nodes. This picture obviously calls for a "repair" of the solution. For instance, replacing the edges from 12 to 28 and from 11 to 29 by the edges from 11 to 12 and 28 to 29 results in considerable saving. Obviously, further improvements of this kind are possible.

The 2-OPT heuristic formalizes this idea. It starts with some tour T, for instance a random tour or a tour obtained by a construction heuristic. Then it checks, for all pairs of non-adjacent edges uv, xy of T, whether the unique tour S formed by deleting these two edges and adding two edges, say $S := T \setminus \{uv, xy\} \cup \{ux, vy\}$ is shorter than T (this is the case, e.g., when the segments uv and xy cross in the plane). If so, T is replaced by S and the exchange tests are repeated. Otherwise the heuristic stops. Figure 3.1 shows the result of 2-OPT started with the tour in fig. 2.1.

There is an obvious generalization of this method: instead of removing two

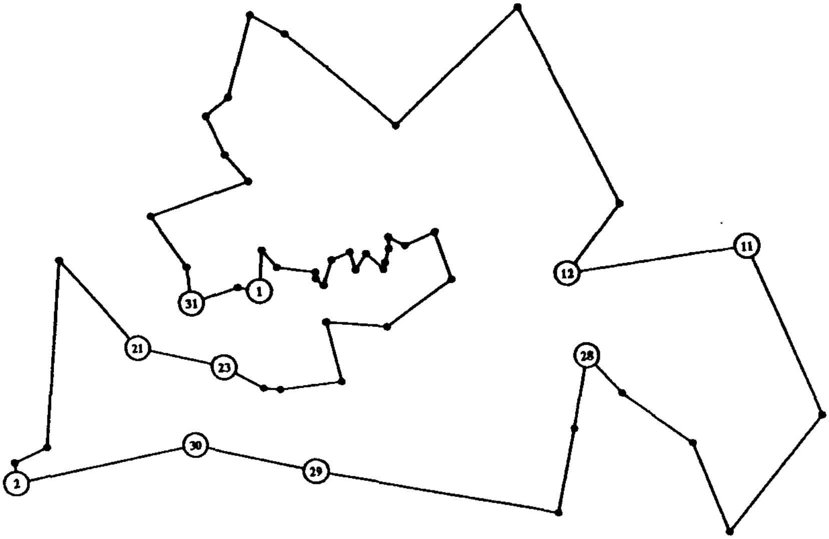


Figure 3.1. A 2-OPT tour of a 52-city-problem.

edges, we delete r non-adjacent edges from T. Then we enumerate all possible ways of adding r other edges such that these r edges together with the remaining r paths form a tour. If the shortest of these tours is shorter than the present tour T we replace T by this shorter tour and repeat. This method is called the r-OPT heuristic.

These heuristics are prototypical *local improvement techniques* that, in general, work as follows. We have some feasible solution of the given combinatorial optimization problem. Then we do some little operations on this solution, such as removing some elements and adding other elements, to obtain one or several new solutions. If one of the new solutions is better than the present one, we replace the present one by the new best solution and repeat.

The basic ingredient of such improvement or exchange heuristics is a rule that describes the possible manipulations that are allowed. This rule implicitly defines, for every feasible solution, a set of other feasible solutions that can be obtained by such a manipulation. Using this interpretation, we can define a digraph D whose vertex set is the set Ω of feasible solutions of our combinatorial optimization problem (this is typically exponentially large) and where an arc (S, T) is present if T can be obtained from S by the manipulation rule.

Now, local improvement heuristics can be viewed as algorithms that start at some node of this digraph and search, for a current node, the successors of the node. If they find a better successor they go to this successor and repeat. The term *local search algorithms* that is also used for these techniques derives from this interpretation.

There are many important algorithms that fit in this scheme: the simplex method (see chapter 30), basis reduction algorithms (chapter 20), etc. The simplex method is somewhat special in the sense that it only gets stuck when we have reached the optimum. Usually, local improvement algorithms may run into *local optima*, i.e., solutions from which the given local manipulations do not lead to any better solution.

Maximum cuts. Let us look at another example. Suppose we are given a graph $G=(V,E)$ and we want to find a cut $\delta(W)=\{ij\in E: i\in W,\ j\not\in W\}$, $W\subseteq V$, of maximum size. (This is *the cardinality max-cut problem* for G.) The problem has a natural weighted version, where each edge has a (say, rational) weight, and we want to find a cut with maximum weight.

We start with an arbitrary subset $W\subseteq V$. We check whether W (or $V\backslash W$) contains a node w such that fewer than half of its neighbors are in $V\backslash W$) (or W). If such a node exists we move it from W to $V\backslash W$ (or from $V\backslash W$ to W). Otherwise we stop.

Termination of this *single exchange heuristic* in $O(n^2)$ time is guaranteed, since the size of the cut increases in every step. The cut produced by this procedure obviously has a size that is at least one half of the maximum cardinality of a cut in G, and in fact, it can be as bad as this, as the complete bipartite graph shows.

The single exchange heuristic has an obvious weighted version: we push a node w to the other side if the sum of the weights of edges linking w to nodes on the other side is smaller than the sum of the weights of the edges linking w to nodes on the side of w. This version, of course, also terminates in finite time, but the number of steps may be exponential (Haken and Luby 1988) even for 4-regular graphs. On the other hand Poljak (1993) proved that the single exchange heuristic for the weighted max-cut problem terminates in polynomial time for cubic graphs (see section 9).

For a typical combinatorial optimization problem, it is easy to find many local manipulation techniques. For instance, for the max-cut problem we could try to move several nodes from one side to the other or exchange nodes between sides; for the TSP we could perform node exchanges instead of edge exchanges, we could exchange whole sections of a tour, and we could combine these techniques, or we could vary the number of edges and/or nodes that we exchange based on some criteria. In fact, most people working on the TSP view the well-known heuristic of Lin and Kernighan (1973) and its variants as the best local improvement heuristics for the TSP known to date (based on the practical performance of heuristics on large numbers of TSP test instances). This heuristic is a dynamic version of the r-OPT heuristic with varying r.

The last statement may suggest that improvement heuristics are the way to solve (large scale) TSP instances approximately. However, there are some practical and theoretical difficulties with respect to running times, traps and worst-case behavior.

Running time of exchange heuristics. A straightforward implementation of the r-OPT heuristic, for example, is hopelessly slow. It takes $\binom{n}{r}$ tests to check whether a tour can be improved by an r-exchange. Even for $r = 3$, the heuristic runs almost forever on medium size instances of a few thousand nodes. To make this approach practical, a number of modifications limiting the exchanges considered are necessary. They are usually based on insights about the probability of success of certain exchanges, or on knowledge about special structures (an instance might be geometrical, e.g., given by points in the plane and a distance function). Well-designed fast data structures play an important role. The issue of speeding up TSP heuristics is treated in depth, e.g., in Johnson (1990), Bentley (1992), and Reinelt (1994). With these techniques TSP instances of up to a million cities have been solved approximately. Such observations apply to many other combinatorial optimization problems analogously.

It may sound strange, but for many exchange heuristics there is no proof that these heuristics terminate in polynomial time. This is the case even with such a basic and classical algorithm as the simplex method! For certain natural pivoting rules we know that they may lead to exponentially many iterations, while for others, it is not known whether or not they terminate in polynomial time; no pivoting rule is known to terminate in polynomial time. As another, simpler example, we mention that although we could prove an $O(n^2)$ running time for the cardinality version of the max-cut heuristic, its weighted version is not polynomial, as mentioned above.

A single pass through the loop of the r-OPT heuristic for the TSP takes $O(n^r)$ time but it is not clear how to bound the number of tours that have to be processed before the algorithm terminates with an r-OPT *tour*, i.e., a tour that cannot be improved by an r-exchange. Computational experience, however, shows that exchange heuristics usually do not have to inspect too many tours until a "local optimum" is found.

Quality of the approximation. Although the solution quality of local improvement heuristics is often quite good, these heuristics may run into traps, for instance r-OPT tours, whose value is not even close to the optimum. It is, in general, rather difficult to prove worst-case bounds on the quality of exchange heuristics. For an example where a performance ratio is established, see the basis reduction algorithm in chapter 19.

It is probably fair to say that, for the solution of combinatorial optimization problems appearing in practice, fast construction heuristics combined with local improvement techniques particularly designed for the special structures of the application are the real workhorses of combinatorial optimization. That is why this machinery receives so much attention in the literature and why new little tricks or clever combinations of old tricks are discussed intensively. Better solution qualities or faster solution times may result in significant cost savings, in particular for complicated problems of large scale.

Aiming at local optima. There are several examples when we are only interested in finding a local optimum: it is the structure of the local optimum, and not the value of the objective function, that concerns us. A theoretical framework for such "polynomial local search problems" was developed by Johnson et al. (1988). Analogously to NP, this class also has complete (hardest) problems; the weighted local max-cut problem is one of these (Schäffer and Yannakakis 1991).

Consider an optimum solution W of the max-cut problem for a graph $G = (V, E)$. Clearly, every node is connected to at least as many points on the opposite side of the cut than on its own side. If we only want a cut with this property, any locally optimal cut (with respect to the single exchange heuristic) would do.

Assume now that we want to solve the following more general problem: given two functions $f, g: V \rightarrow \mathbb{Z}_+$ such that $f(v) + g(v) = d_G(v) - 1$ for all $v \in V$, find a subset $W \subseteq V$ such that for each node v, the number of nodes adjacent to v on its own side of the cut is at most

$$\begin{aligned} &f(v)\,, \quad \text{if } v \in W\,, \\ &g(v)\,, \quad \text{if } v \in V \setminus W\,. \end{aligned}$$

It is not difficult to guess an objective function (over cuts) for which such cuts are exactly local optima:

$$\phi(W) := |\delta(W)| + f(W) + g(V \setminus W)\,.$$

Once this is found, it follows that a cut with the desired property exists, and also that it can be found in polynomial time by local improvement. A much more difficult, but in principle similar, application of this idea is the proof of Szemerédi's regularity lemma (see chapter 23). Here again a tricky (quadratic) objective function is set up, which is locally improved until a locally almost optimal solution is found; the structure of such a solution is what is needed in the numerous applications of the Regularity lemma (see also Alon et al. 1994 for the algorithmic aspects of this procedure).

A beautiful example of turning a structural question into an optimization problem is Tutte's (1963) proof of the fact that every 3-connected planar graph has a planar embedding with straight edges and convex faces. He considers the edges as rubber bands, fixes the vertices of one face at the vertices of a convex polygon, and lets the remaining vertices find their equilibrium. This means minimizing a certain quadratic objective function (the energy), and the optimality criteria can be used to prove that this equilibrium state defines a planar embedding with the right properties. The algorithm can be used to actually compute nice embeddings of planar graphs. A similar method for connectivity testing was given by Linial et al. (1988).

Randomized exchange. A very helpful idea to overcome the problem of falling into a trap is to randomize. In a randomized version of local search, a random neighbor of the current feasible solution is selected. If this improves the objective function, the current feasible solution is replaced by this neighbor. If not, it may still be replaced, but only with some probability less than 1, depending on how much worse the objective value at the neighbor is. This relaxation of the strict descent rule may help to jump out of traps and eventually reach a significantly better solution.

A more general way of looking at this method is to consider it as generating a random element in the set Ω of feasible solutions, from some given probability distribution Q. Let $f: \Omega \rightarrow \mathbb{R}_+$ be the objective function; then maximizing f over Ω is just the extreme case when we want to generate a random element from a distribution concentrated on the set of optimum solutions. If, instead, we generate a random point w from the distribution Q in which $Q(v)$ is proportional to $\exp(f(v)/T)$, where T is a very small positive number, then with large probability w will maximize f. In fact, a randomized algorithm that finds a solution that is (nearly) optimal with large probability is equivalent to a procedure of generating a random element from a distribution that is heavily concentrated on the (nearly) optimal solutions.

To generate a random element from a distribution over a (large and complicated) set is of course a much more general question, and is a major ingredient in various algorithms for enumeration, integration, volume computation, simulation, statistical sampling, etc. (see Jerrum et al. 1986, Dyer and Frieze 1992, Sinclair and Jerrum 1988, Dyer et al. 1991, Lovász and Simonovits 1992 for some of the applications with combinatorial flavor). An efficient general technique here is *random walks* or *Markov chains*. Let $G = (\Omega, E)$ be a connected graph on Ω, and assume, for simplicity of presentation, that G is non-bipartite and d-regular. If we start a random walk on G and follow it long enough, then the current point will be almost uniformly distributed over Ω. How many steps does "long enough" mean depends on the spectrum, or in combinatorial terms, on global connectivity properties called expansion rate or conductance, of the graph (see also chapter 31).

The Metropolis filter. In optimization, we are interested in very non-uniform, rather than uniform, distributions. Fortunately, there is an elegant way, called the *Metropolis filter* (Metropolis et al. 1953), to modify the random walk, so that it gives any arbitrary prescribed probability distribution. Let $F: \Omega \rightarrow \mathbb{R}_+$. Assume that we are at node v. We choose a random neighbor u. If $F(u) \geq F(v)$ then we move to u; else, we flip a biased coin and move to u only with probability $F(u)/F(v)$, and stay at v with probability $1 - F(u)/F(v)$.

Let Q_F denote the probability distribution on Ω defined by the property that $Q_F(v)$ is proportional to $F(v)$. The miraculous property of the Metropolis filter is the following.

Theorem 3.1. *The stationary distribution of the Metropolis-filtered random walk is* Q_F.

So choosing $F(v) = \exp(f(v)/T)$, we get a randomized optimization algorithm.

Unfortunately, the issue of how long one has to walk gets rather messy. The techniques to estimate the conductance of a Metropolis-filtered walk are not general enough, although Applegate and Kannan (1990) have been able to apply this technique to volume computation.

Simulated annealing. Coming back to optimization, let us follow Kirckpatrick et al. (1983), and call the elements of Ω *states* (of some physical system), $1/F(v)$ the *energy* of the state, and T the *temperature*. In this language, we want to find a state with minimum (or almost minimum) energy. A Metropolis-filtered random walk means letting the system get into a stationary state at the given temperature.

The main, and not quite understood, issue is the choice of the temperature. If we choose T large, the quality of the solution is poor, i.e., the probability that it is close to being optimal is small. If we choose T small, then there will be barriers of very small probability (or, equivalently, with very large energy) between local optima, and it will take extremely long to get away from a local, but not global optimum.

The technique of *simulated annealing* suggests to start with the temperature T sufficiently large, so that the random walk with this parameter mixes fast. Then we decrease T gradually. In each phase, the random walk starts from a distribution which is already close to the limiting distribution, so there is hope that the walk will mix fast. (A similar trick works quite well in integration and volume computation; see Lovász and Simonovits 1992.) Theoretical and practical experiments have revealed that it matters a lot how long we walk in a given phase (*cooling schedule*).

There are many empirical studies with this method; see Johnson et al. (1989, 1991) or Johnson (1990). There are also some general estimates on its performance (Holley and Stroock 1988, Holley et al. 1989). Examples of problems, in particular of the matching problem, are known where simulated annealing performs badly (Sasaki and Hajek 1988, Sasaki 1991, Jerrum 1992), and some positive results in the case of the matching problem are also known (Jerrum and Sinclair 1989). The conclusion that can be drawn at the moment is that simulated annealing is a potentially valuable tool (if one can find good cooling schedules and other parameters), but it is in no way a panacea as was claimed in some papers pioneering this topic.

Approaches named taboo search, threshold accept, evolution or genetic algorithms and others are further variants and enhancements of randomized local search. The above judgement of simulated annealing applies to them as well, see Johnson (1990).

4. Enumeration and branch-and-bound

There is a number of interesting combinatorial optimization problems for which beautiful polynomial time algorithms exist. We will explain some of them in subsequent sections. We now address the issue of finding an optimum solution for an NP-hard problem. In the previous two sections we have outlined heuristics that produce some feasible and hopefully good solutions. Such a solution may even be optimal right away. But how does one verify that?

The basic trouble with integer programming and combinatorial optimization is the non-existence of a sensible duality theory. The duality theorem of linear programming (see chapter 30), for instance, can be used to prove that some given feasible solution is optimal. In the (rare) cases where duality theorems in integral solutions like the max-flow min-cut theorem (see chapter 2 or 30) or the Lucchesi–Younger theorem (see chapter 2) exist, one can usually derive a polynomial time solution algorithm. For NP-hard problems one should not expect to find such theorems. Unfortunately, nothing better is known than replacing such a theory by brute force.

Trivial running time estimates reveal that the obvious idea of simply enumerating the finitely many solutions of a combinatorial optimization problem is completely impractical in general. For instance, computing the length of all $\frac{1}{2} \cdot 15!$ ($\sim 0.6 * 10^{12}$) tours of a 16 city TSP instance takes about 90 hours on a 30 MIPS workstation. Even a teraflop computer will be unable to enumerate all solutions of a ridiculously small 30 city TSP instance within its lifetime.

Unless P = NP, there is no hope that we will be able to design algorithms for NP-hard problems that are asymptotically much better than enumeration. However, we can try to bring problem instances of reasonable sizes (appearing in practice, say) into the realm of practical computability by enhancing enumeration with a few helpful ideas.

The idea of the *branch-and-bound* approach is to compute tight upper and lower bounds on the optimum value in order to significantly reduce the number of enumerative steps. To be more specific, let us assume that we have an instance of a minimization problem. Let Ω be its set of feasible solutions.

To implement branch-and-bound, we need a dual heuristic (relaxation), i.e., an efficiently computable lower bound on the optimum value. This dual heuristic will also be called for certain subproblems.

We first run some construction and improvement heuristics to obtain a good feasible solution, say T, with value $c(T)$, which is an upper bound for the optimum value c_{opt}.

Now we resort to enumeration. We split the problem into two (or more) subproblems. Recursively solving these subproblems would mean straightforward enumeration. We can gain by maintaining the best solution found so far and computing, whenever we have a subproblem, a lower bound for the optimum value of this subproblem, using the dual heuristic. If this value is larger than the value of the best solution found so far, we do not have to solve this subproblem.

To be more specific, let us discuss a bit the two main ingredients, *branching* and *bounding*.

We assume that the splitting into subproblems (the *branching*) is such that the set Ω of feasible solutions is partitioned into the sets Ω' and Ω'' of feasible solutions of the subproblems. We also assume that the subproblems are of the same type (e.g., the dual heuristic applies to them). It is also important that the branching step requires little bookkeeping and is computationally cheap. If Ω

consists of subsets of a set S, then a typical split is to choose an element $e \in S$ and to set

$$\Omega' := \{I \in \Omega \mid e \in I\}\,, \qquad \Omega'' := \{I \in \Omega \mid e \not\in I\}\,.$$

The *bounding* is usually provided by a *relaxation*: by problem-specific investigations we introduce a new problem, whose set of feasible solutions is Γ, say, such that $\Omega \subseteq \Gamma$ and the objective function for Ω extends to Γ. Suppose X minimizes the objective function over Γ, then the value $c(X)$ provides a lower bound for c_{opt} since all elements of Ω participated in the minimization process. If X is, in fact, an element of Ω we clearly have found an optimum solution of Ω. (If $X \not\in \Omega$, we may still be able to make use of it by applying a construction and/or improvement heuristic that starts with X and ends with a solution Y, say, in Ω. If $c(Y) < c(T)$ we set $T := Y$ to keep track of our current best solution.)

To give some examples, useful relaxations for the symmetric TSP are perfect 2-matchings (unions of disjoint circuits that cover all nodes) or 1-trees (cf. sections 2 and 9). For the asymmetric TSP a standard relaxation is obtained by considering unions of directed circuits that cover all nodes (which can be easily reduced to a bipartite perfect matching problem), or the r-arborescence problem.

A particularly powerful method is based on LP-relaxations. This is covered in depth in section 8. There are some general methods to improve relaxations; one technique is called Lagrangian relaxation and will be discussed in section 9.

Returning to the algorithm, we maintain a list of unsolved subproblems, and a solution T that is the current best. In the general step, we choose an unsolved subproblem, say Ω^i, from the list and remove it. We optimize the objective function over the relaxation Γ^i of Ω^i. Let X be an optimum solution. There are several possibilities.

- First, X may be feasible for Ω^i. In this case we have found an optimum over Ω^i and can completely eliminate all elements of Ω^i from the enumeration process. One often says that this branch is *fathomed*. If $c(X) < c(T)$, we reset $T := X$ also.
- Second, if $c(X) \geq c(T)$ then no solution in Γ^i and hence no solution in Ω^i has a value that is smaller than the current champion. Hence this branch is also fathomed and we can eliminate all solutions in Ω^i.
- Third, if $c(X) < c(T)$ (and X is not feasible for Ω^i), we have done the computation in vain. (We may still try to make use of X to obtain a solution better than T as above.) We split Ω^i into two or more pieces, and put these on our list of unsolved subproblems.

The branch-and-bound method terminates when the list of subproblems is empty. The iteratively updated solution T is the optimum solution. Termination is, of course, guaranteed if the set Ω of feasible solutions is finite.

Although the global procedure is mathematically trivial, it is a considerable piece of work to make it computationally effective. The efficiency mainly depends on the quality of the lower bound used. Most of the mathematics that is developed for the solution of hard problems is concerned with the invention of better and better relaxations, with their structural properties and with fast algorithms for their solution.

5. Dynamic programming

Dynamic programming is a general technique for optimum decision making. It was originally developed (Bellman 1957) for the solution of discrete-time sequential decision processes. The process starts at a given initial state. At any time of the process we are in some state and there is a set of states that are reachable from the present state. We have to choose one of these. Every state has a value and our objective is to maximize the value of the terminating state. Such an optimization problem is called a *dynamic program*.

Virtually any optimization problem can be modeled by a dynamic program. There is a recursive solution for dynamic programs which, however, is not efficient in general. But the dynamic programming model and this recursion can be used to design fast algorithms in cases where the number of states can be controlled. We illustrate this by means of a few examples.

The subset-sum problem. Given positive integers $a_1, a_2, \ldots, a_n, b$, decide whether there exist indices $1 \leq i_1 < i_2 < \cdots < i_k \leq n$ for some k such that $a_{i_1} + \cdots + a_{i_k} = b$. This problem, called *subset-sum problem*, is NP-complete; however, there is a pseudopolynomial algorithm to solve it.

First, consider an obvious algorithm using enumeration. Clearly the subset-sum problem has a solution for a given input $(a_1, \ldots, a_n, b)$, if and only if it has a solution either for $(a_1, \ldots, a_{n-1}, b - a_n)$ or for $(a_1, \ldots, a_{n-1}, b)$. So an instance of the subset-sum problem for n numbers can be reduced to two subproblems with $n-1$ numbers each. Building up a search tree based on this observation yields an $O(2^n)$ algorithm (which basically enumerates all subsets of the a_i).

But looking at this tree more carefully, we see that, at least if b is small compared with 2^n, it has a crucial property: *the same subproblem occurs on many branches*! In fact, there are only nb distinct subproblems altogether: for each $c \leq b$ and each $m \leq n$, the subset-sum problem with input $a_1, \ldots, a_m, c$. Imagine that the branches of the search tree "grow together" if the same subproblem occurs: we get a "search diagraph" D. The nodes of this acyclic digraph are labelled with pairs (c, m), and there is an edge from (c, m) to $(c', m-1)$ if either $c' = c$ or $c' = c - a_m$. The subset-sum problem is solvable if and only if there is a dipath from (b, n) to $(0, 0)$. Such a dipath can be found (if it exists) in $O(bn)$ time by searching D either from $(0, 0)$ or from (b, n).

Along the same lines, one can devise an algorithm for the knapsack problem with running time polynomial in $b + \sum_i \langle c_i \rangle$.

Minimal triangulation of a convex polygon. Given a convex polygon P with n

vertices in the plane, we want to find a triangulation with minimal total edge length. (The length c_{ij} of each edge ij is known.)

If the vertices of P are numbered consecutively 1 through n, take edge $1n$ and consider the vertex i with which it forms a triangle in the triangulation. For a given i, it suffices to find optimal triangulations of the two polygons with vertices $1, \ldots, i$ and $i, \ldots, n$, respectively, which can be done independently, see fig. 5.1. So we have produced $2(n-2)$ subproblems.

If we are not careful, repeating this process could lead to exponentially many distinct subproblems. But note that if we choose the triangle containing the edge $1i$ to cut the polygon $1, \ldots, i$, then we get two subproblems corresponding to convex polygons having only one edge that is not an edge of P.

In general, given two vertices i and j with $i \leqslant j-2$, let $f(i, j)$ denote the minimum total length of diagonals triangulating the polygon with vertices $(i, i+1, \ldots, j)$. Then clearly $f(i, i+2) = 0$ and

$$f(i, j) = \min\{\min_{i+2 \leqslant k \leqslant j-2} \{f(i,k) + f(k,j) + c_{ik} + c_{kj}\},\; f(i+1, j) + c_{i+1,j},\, f(i, j-1) + c_{i,j-1}\}\,. \tag{4}$$

The answer to the original question is $f(1, n)$.

We can represent the computation by a "search digraph" whose nodes correspond to all the polygons with vertex set $(i, i+1, \ldots, j)$, where $1 \leqslant i, j \leqslant n$, $i \leqslant j-2$. We set $f(i, j) = 0$ if $j = i+2$, and can use (4) recursively if $j > i+2$. There are $O(n^2)$ subproblems to solve, and each recursive step takes $O(n)$ time. So we get an $O(n^3)$ algorithm.

Steiner trees in planar graphs. Let $G = (V, E)$ be a graph with edge lengths $c_e > 0$, and let $T \subseteq V$ be a set of "terminals". A *Steiner tree* in G is a subtree of G that contains all nodes of T. The *Steiner tree problem* is the task of finding a shortest

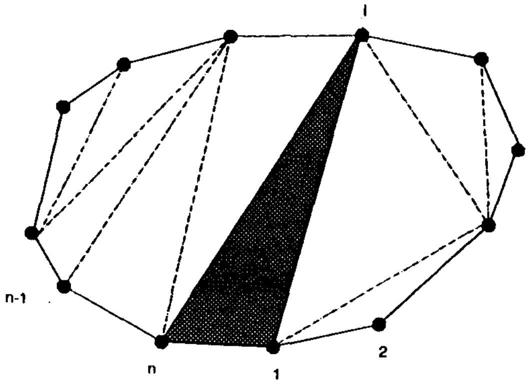


Figure 5.1. Optimum triangulation of a convex polygon.

Steiner tree. This problem is NP-hard in general, even for planar graphs. But in the case of a planar graph when all the terminal nodes are on one, say, on the outer face C, a shortest Steiner tree can be found by dynamic programming as follows.

Let us first look at a minimum Steiner tree B. Pick any node v of B and let B' be the union of some branches of B that are rooted at v and that are, in addition, consecutive in the natural cyclic order of the edges leaving v. Let $T' := T \cap V(B')$. We observe the following (see fig. 5.2, where v is represented by a black circle):

- There is a path $P \subseteq C$ whose endnodes are terminals such that $T' = V(P) \cap T$.
- B' is a minimum length Steiner tree with respect to the terminal set $T' \cup \{v\}$.
- v is on the outer face of the subgraph $B' \cup P$.

These observations motivate the following dynamic program for the solution of our Steiner tree problem. For every path $P \subseteq C$ whose end nodes are terminals and every node $v \in V$, we determine a shortest Steiner tree B' with respect to the set of terminal nodes $(V(P) \cap T) \cup \{v\}$ with the additional requirement that v is on the outer face of $B' \cup P$.

If P consists of just one terminal then such a Steiner tree can be found by a shortest path calculation.

Suppose that we have solved this subproblem for all nodes $v \in V$ and all paths $P \subseteq C$ containing at most k terminal nodes. To solve the subproblem for some node $v \in V$ and a path $P \subseteq C$ containing $k+1$ terminal nodes, we do the

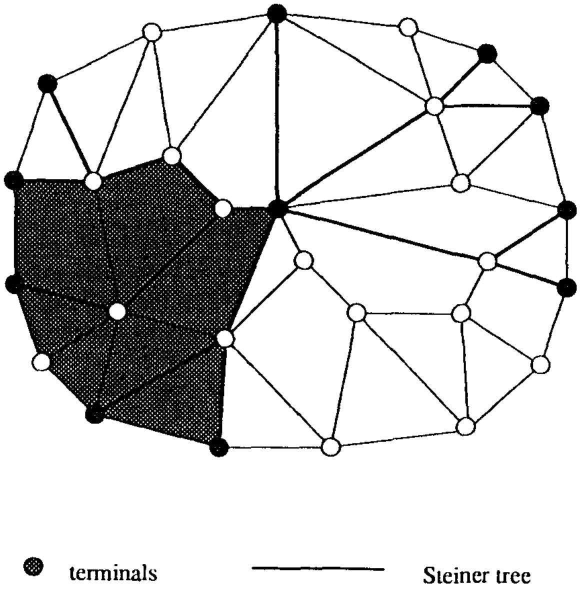


Figure 5.2. Steiner tree in planar graphs.

following. Let $t_1, \ldots, t_{k+1}$ be the terminals contained in P in the natural order. For every node $w \in V$ and every two subpaths P_1, P_2 of P, where P_1 connects t_1 to t_j and P_2 connects t_{j+1} to t_{k+1}, $1 \leq j \leq k$, we solve the subproblems for w and P_1 and for w and P_2 to get two trees B_1 and B_2. We also compute a shortest path Q from w to v. Among all the sets $B_1 \cup B_2 \cup Q$ computed this way we choose the one with minimum length. This is an optimum solution of our subproblem for v and P.

To get a minimum length Steiner tree, consider a path $P \subseteq C$ that contains all terminal nodes and choose the shortest among all solutions of subproblems for v and P with $v \in V$.

This algorithm is due to Erickson et al. (1987) and is based on ideas of Dreyfus and Wagner. It can be extended to the case when all terminals are on a fixed number of faces.

There are many other non-trivial applications of the idea of dynamic programming; for example, Chvátal and Klincsek (1980) use it to design a polynomial time algorithm that finds a maximum cardinality subset of a set of n points in the plane that forms the vertices of a convex polygon (cf. chapter 17, section 7.2).

Optimization on tree-like graphs. There are many NP-hard problems that are easy if the underlying graph is a tree. Consider the stable set problem in a tree T. We fix a root r and, for every node x, we consider the subtree T_x consisting of x and its descendants. Starting with the leaves, we compute, for each node x, two numbers: the maximum number of independent nodes in T_x and in $T_x - x$. If these numbers are available for every son of x, then it takes only $O(d(x))$ time to find them for x. Once we know them for T_r, we are done. So a maximum stable set can be found in a linear time.

Similar algorithms can be designed for more general "tree-like" graphs, e.g., series–parallel graphs. A general framework for "tree-like" decompositions, developed by Robertson and Seymour in their "Graph minors" theory, leads to very general dynamic programming algorithms on graphs with bounded tree-width. See chapter 5 for the definition of tree-width and for examples of such algorithms.

6. Augmenting paths

In local search algorithms, we try to find very simple "local" improvements on the current solution. There are more sophisticated improvement techniques that change the current solution globally, usually along paths or systems of paths. These methods are often called *augmenting path* techniques. Since they occur throughout this Handbook, we refrain from describing any of them here; let it suffice to quote the most important applications of the method of alternating paths: maximum flows and packing of paths, chapter 2; maximum matchings (weighted and unweighted), maximum stable sets in claw-free graphs, chapter 3;

edge-coloring, chapter 4; matroid intrersection matroid matching, and submodular flows, chapter 11.

7. Uncrossing

Uncrossing is a technique to simplify complicated set-systems, while maintaining certain key properties. One can find many applications of the uncrossing procedure in this Handbook as a proof technique; it is applied in the theory of graph connectivity and flows (chapter 2), matchings (chapter 3, or Lovász and Plummer 1986), and matroids (chapters 11, 30). It is worth pointing out, however, that uncrossing can be viewed as an algorithmic tool, that constructs, from a complicated dual solution, a dual solution with a tree-like structure. This way it is sometimes possible to derive an optimum integral dual solution from an optimum fractional dual solution.

As an illustration, consider the problem of finding a maximum family of rooted cuts in a digraph $G = (V, A)$ with root r such that every arc a occurs in at most l_a of these cuts, where the $l_a \geqslant 0$ are given integer values ("lengths"). (A rooted cut or r-cut, is the set of arcs entering S, i.e., with tail in $V \setminus S$ and head in S for some non-empty $S \subset V$, $r \not\in S$; cf. chapter 30.) Assume that we have a *fractional packing*, i.e., a family $\mathscr{F}$ of r-cuts and a weight $w_D \geqslant 0$ for every $D \in \mathscr{F}$ such that $\sum_{D \ni e} w_D \leqslant l_a$ for every arc a (ellipsoidal or interior point methods, as well as averaging procedures, may yield such "fractional solutions"). As a consequence of Fulkerson's optimum arborescence theorem (chapter 30, Theorem 5.7), we know that there exists an integer solution with the same value, i.e., a family of at least $\sum_D w_D$ r-cuts with the prescribed property. But how to find this?

For each r-cut D in the digraph G, we denote by $S(D)$ a set $S \subseteq V \setminus \{r\}$ such that D is the set of edges entering S. Let $\mathscr{H} = \{S(D){:}\ D \in \mathscr{F}\}$. Call two r-cuts D_1 and D_2 *intersecting* if all three sets $S(D_1) \cap S(D_2)$, $S(D_1) \setminus S(D_2)$ and $S(D_2) \setminus S(D_1)$ are non-empty. Assume that $\mathscr{F}$ contains two intersecting cuts D_1 and D_2, and let D' and D'' denote the r-cuts defined by $S(D_1) \cap S(D_2)$ and $S(D_1) \cup S(D_2)$, respectively.

Decrease w_{D_1} and w_{D_2} by ε and increase $w_{D'}$ and $w_{D''}$ by ε, where $\varepsilon := \min\{w_{D_1}, w_{D_2}\}$ (if, say, D' does not belong to $\mathscr{F}$, then we add it to $\mathscr{F}$ with $w_{D'} = 0$). It is easy to check that this yields a new fractional packing with the same total weight. The family $\mathscr{F}$ lost one member (one of D_1 and D_2), and gained at most two new members (D' and D''). If the new family contains two intersecting cuts, then we "uncross" them as above. It can be shown that the procedure terminates in a polynomial number of steps (see Hurkens et al. 1988 for a discussion of this).

When the uncrossing procedure terminates, the family $\mathscr{H}$ is *nested*, i.e., there are no intersecting pairs of cuts in $\mathscr{F}$. Such a family has a tree structure; $\mathscr{H}$ can be obtained by selecting disjoint subsets of $V \setminus \{r\}$, then disjoint subsets in these subsets, etc. It is not difficult to see that the number of members of $\mathscr{H}$ is at most $2|V| - 3$.

Choose $D \in \mathscr{F}$ such that w_D is not an integer and $S(D)$ is minimal. There is a unique cut $D' \in \mathscr{F}$ such that $S(D') \supset S(D)$ and $S(D')$ is minimal. Add ε to w_D and subtract ε from $w_{D'}$, where $\varepsilon := \min\{\lceil w_D \rceil - w_D, w_{D'}\}$. It is easy to check using the integrality of ℓ) that this results in a fractional packing with the same value, and now either w_D is an integer or $w_{D'} = 0$. After at most $2n - 3$ repetitions of this shift, we get a fractional packing with all weights integral, which trivially gives the family as required.

8. Linear programming methods

A very successful way to solve combinatorial optimization problems is to translate them into optimization problems for polyhedra and utilize linear programming techniques. The theoretical background of this approach is surveyed in chapter 30 where also many examples of the application of this method are provided. We will concentrate here on the implementation of the linear programming approach to practical problem solving, and on the use of linear programming in heuristics.

We will assume that we have a combinatorial optimization problem with linear objective function like the travelling salesman, the max-cut, the stable set, or the matching problem. Let us also assume that we want to find a feasible solution of maximum weight. Typically, an instance of such a problem is given by a ground set E, an objective function $c: E \to \mathbb{R}$ and a set $\mathscr{I} \subseteq 2^E$ of feasible solutions such as the set of tours, of cuts, of stable sets, or matchings of a graph. We transform $\mathscr{I}$ into a set of geometric objects by defining, for each $I \in \mathscr{I}$, a vector $\chi^I \in \mathbb{R}^E$ with $\chi^I_e = 1$ if $e \in I$ and $\chi^I_e = 0$ if $e \not\in I$. The vector χ^I is called the *incidence* (*or characteristic*) *vector* of I. Now we set

$$\mathscr{P}(\mathscr{I}) := \operatorname{conv}\{\chi^I \in \mathbb{R}^E \mid I \in \mathscr{I}\} ,$$

i.e., we consider a polytope whose vertices are precisely the incidence vectors of the feasible solutions of our problem instance. Solving our combinatiorial optimization problem is thus equivalent to finding an optimum vertex solution for the following linear program

$$\max c^{\mathrm{T}}x , \quad x \in P(\mathscr{I}) . \tag{5}$$

However, (5) is only a linear program "in principle" since the usual LP-codes require the polyhedra to be given by a system of linear equations and inequalities. Classical results of Weyl and Minkowski ensure that a set given as the convex hull of finitely many points has a representation by means of linear equations and inequalities (and vice versa). But it is by no means simple to find, for a polyhedron given in one of these representations, a complete description in the other way. By problem specific investigations one can often find classes of valid and even facet-defining inequalities that partially describe the polyhedra of interest. (Many of the known examples are described in chapter 30.)

What does "finding" a class mean? The typical situation in polyhedral combinatorics is the following. A class of inequalities contains a number of inequalities that is exponential in $|E|$. It is well-described in the sense that we can (at least) decide in polynomial time whether a given inequality, for a given instance, belongs to the class. This is a minimal requirement; we need more for a class to be really useful (see *separation* below). Also, the class should contain strong inequalities; best is when most of the inequalities define facets of $P(\mathcal{I})$. Except for special cases, one such class (or even a bounded number of such classes) will not provide a complete description, i.e., $P(\mathcal{I})$ is strictly contained in the set of solutions satisfied by all these inequalities.

The cut polytope. To give an example, let us discuss the max-cut problem. In this case a graph $G = (V, E)$ is given (for convenience we will assume that G is simple), and we are interested in the convex hull of all incidence vectors of cuts in G, i.e.,

$$\mathrm{CUT}(G) := \mathrm{conv}\{\chi^{\delta(W)} \in \mathbb{R}^E \mid W \subseteq V\} .$$

This polytope has dimension $|E|$. For any edge $e \in E$, the two *trivial inequalities* $0 \leqslant x_e \leqslant 1$ define a facet of CUT(G) if and only if e is not contained in a triangle. About some other classes of facets, we quote the following results of Barahona and Mahjoub (1986).

Theorem 8.1. *Let $G = (V, E)$ be a graph.*

(a) *For every cycle $C \subseteq E$ and every set $F \subseteq C$, $|F|$ odd, the* odd cycle inequality

$$x(F) - x(C)\backslash F) := \sum_{e \in F} x_e - \sum_{e \in C \backslash F} x_e \leqslant |F| - 1$$

is valid for CUT(G); *it defines a facet of* CUT(G) *if and only if C has no chord.*

(b) *For every complete subgraph $K_p = (W, F)$ of G, the* K_p-inequality

$$x(F) \leqslant \left[\frac{p}{2}\right]\left[\frac{p}{2}\right]$$

is valid for CUT(G); *it defines a facet of* CUT(G) *if and only if p is odd.*

Applications of the max-cut problem arise in statistical mechanics (finding ground states of spin glasses, see chapter 37) and VLSI design (via minimization). Both applications are covered in Barahona et al. (1988). But the max-cut problem comes up also in many other fields. Structural insights from different angles resulted in the discovery of many further (and very large) classes of valid and facet-defining inequalities. Studies on the embeddability of finite metric spaces, for instance, lead to the class of hypermetric inequalities; there are the classes of clique-web, suspended tree, circulant, path-block-cycle, and other inequalities. A comprehensive survey of this line of research can be found in Deza and Laurent (1991); cf. also chapter 41, section 2.

There are a few special cases where it is known that some of the classes of inequalities suffice for a characterization of CUT(G). For example, setting

$$P_C(G) := \{x \in \mathbb{R}^E \mid 0 \leqslant x_e \leqslant 1 \text{ for all } e \in E,\ x(F) - x(C \setminus F) \leqslant |F| - 1 \text{ for all cycles } C \subseteq E \text{ and all } F \subseteq C,\ |F| \text{ odd}\}\,, \tag{6}$$

Barahona and Mahjoub showed that CUT(G) = $P_C(G)$ holds if and only if G is not contractible to the complete graph K_5. But for a general graph G, the union of all the known classes of inequalities does not provide a complete description of CUT(G) at all.

Separation. Let us review at this point what has been achieved by this polyhedral approach. We started out with a polytope $P(\mathcal{I})$ and found classes of inequalities $A_1x \leqslant b_1$, $A_2x \leqslant b_2, \ldots, A_kx \leqslant b_k$, say, such that all inequalities are valid and many facet-defining for $P(\mathcal{I})$. The classes are huge and thus we are unable to use linear programming in the conventional way by inputting all constraints. Moreover, even if we could solve the LPs, it is not clear whether the results provide helpful information for the solution of our combinatorial problem. Although the situation looks rather bad at this point we have done a significant step towards solving hard combinatorial optimization problems in practice. We will now outline why.

A major issue is to figure out how one can solve linear programs of the form

$$\begin{array}{ll} \text{maximize} & c^{\mathrm{T}}x \\ \text{subject to} & A_1x \leqslant b_1 \\ & \vdots \\ & A_kx \leqslant b_k \end{array} \tag{7}$$

where some of the matrices A_i have a number of rows that is exponential in $|E|$, and are only implicitly given to us. To formulate the answer to this question we introduce the following problem.

Separation problem. Let $Ax \leqslant b$ be an inequality system and y a vector, determine whether y satisfies all inequalities, and if not, find an inequality violated by y.

Suppose now that we have a class $\mathcal{A}$ of inequality systems $Ax \leqslant b$. (Example: Consider the class consisting of all odd cycle inequalities for CUT(G), for each graph G.) For each system $Ax \leqslant b$, let $\varphi := \min(\langle a_i \rangle + \langle \beta_i \rangle)$, where the minimum is taken over all rows $a_ix \leqslant \beta_i$ of the system. We say that the *optimization problem for $\mathcal{A}$ can be solved in polynomial time* if, for any system $Ax \leqslant b$ of $\mathcal{A}$ and any vector c, the linear program $\max\{c^{\mathrm{T}}x \mid Ax \leqslant b\}$ can be solved in time polynomial in $\varphi + \langle c \rangle$, and we say that the *separation problem for $\mathcal{A}$ can be solved in polynomial time* if, for any system $Ax \leqslant b$ of $\mathcal{A}$ and any vector y, the separation problem for $Ax \leqslant b$ and y can be solved in time polynomial in φ and $\langle y \rangle$.

Theorem 8.2. *Let $\mathcal{A}$ be a class of inequality systems, then the optimization problem for $\mathcal{A}$ is solvable in polynomial time if and only if the separation problem for $\mathcal{A}$ is solvable in polynomial time.*

For a proof and extensions of this result, see Grötschel et al. (1988).

The idea now is to develop polynomial time separation algorithms for the inequality systems $A_1x \leq b_1, \ldots, A_kx \leq b_k$ in (7). It turns out that this task often gives rise to new and interesting combinatorial problems and that, for many hard combinatorial optimization problems, there are large systems of valid inequalities that can be separated in polynomial time.

We use the max-cut problem again to show how *separation algorithms*, i.e., algorithms that solve the separation problem can be designed. We thus assume that a graph $G = (V, E)$ is given and that we have a vector $y \in \mathbb{R}^E$, $0 \leq y_e \leq 1$ for all $e \in E$. We want to check whether y satisfies the inequalities described in Theorem 8.1.

To solve the *separation problem for the odd cycle inequalities of Theorem* 8.1(a) in polynomial time, we define a new graph $H = (V' \cup V'', E' \cup E'' \cup E''')$ that consists of two copies of G, say $G' = (V', E')$ and $G'' = (V'', E'')$ and the following additional edges E'''. For each edge $uv \in E$ we create the two edges $u'v''$ and $u''v'$. The edges $u'v' \in E'$ and $u''v'' \in E''$ are assigned the weight y_{uv}, while the edges $u'v''$, $u''v' \in E'''$ are assigned the weight $1 - y_{uv}$. For each pair of nodes $u', u'' \in W$, we calculate a shortest (with respect to the weights just defined) path in H. Such a path contains an odd number of edges of E''' and corresponds to a closed walk in G containing u. Clearly, if the shortest of these (u', u'')-paths in H has length less than 1, there exists a cycle $C \subseteq E$ and an edge set $F \subseteq C, |F|$ odd, such that y violates the corresponding odd cycle inequality. (C and F are easily constructed from a shortest path.) If the shortest of these (u', u'')-paths has length at least 1, then y satisfies all these inequalities (see Barahona and Mahjoub 1986).

Trivially, for p fixed, one can check all K_p-inequalities in polynomial time by enumeration, but it is now known whether there is a polynomial time algorithm to solve the separation problem for all complete subgraph inequalities of Theorem 8.1(b). In this case one has to resort to *separation heuristics*, i.e., algorithms that try to produce violated inequalities but that are not guaranteed to find one if one exists.

It is a simple matter to show that the integral vectors in the polytope $P_C(G)$, see (7), are exactly the incidence vectors of the cuts of G. This shows that every integral solution of the linear program

$$\begin{aligned} &\text{maximize} \quad c^T x \\ &\qquad \text{(i)} \quad 0 \leq x \leq 1 \\ &\qquad \text{(ii)} \quad x \text{ satisfies all odd cycle inequalities of Theorem 8.1(a)} \\ &\qquad \text{(iii)} \quad x \text{ satisfies all } K_p\text{-inequalities of Theorem 8.1(b)} \end{aligned} \tag{8}$$

is the incidence vector of a cut of G. In particular, the optimum value of (8) (or any subsystem thereof) provides an upper bound for the maximum weight of a cut.

Theorem 8.2 and the exact separation routine for odd cycle inequalities outlined above show that the linear program (8) (without system (iii)) can be solved in polynomial time. So an LP-relaxation of the max-cut problem can be solved in polynomial time that contains (in general) exponentially many inequalities facet-defining for the cut polytope CUT(G). The question now is whether this technique is practical and whether it will help solve max-cut and other hard combinatorial optimization problems.

Outline of a standard cutting plane algorithm. Theorem 8.2 is based on the ellipsoid method. Although the algorithm that proves 8.2 is polynomial, it is not fast enough for practical problem solving. To make this approach usable in practice one replaces the ellipsoid method by the simplex method and enhances it with a number of additional devices. We will sketch the issues coming up here. We assume that, by theoretical analysis, we have found an LP-relaxation such as (7) of our combinatorial optimization problem.

The initial linear program. We group the inequalities of (7) such that the system $A_1x \leq b_1$ is not too large and contains those inequalities that we feel should be part of our LP initially.

This selection is a matter of choice. In the max-cut problem, for instance, one would clearly select the trivial inequalities $0 \leq x \leq 1$. For the large classes of the other inequalities, the choice is not apparent. One may select some inequalities based on heuristic procedures. In the case of the travelling salesman problem, see (9.10) of chapter 30 and section 2 of this chapter, in addition to the trivial inequalities, the degree constraints $x(\delta(v)) = 2$ for all $v \in V$ are self-suggesting. For the packing problem of Steiner trees (a problem coming up in VLSI routing), for example, a structural analysis of the nets to be routed on the chip was used in Grötschel et al. (1992a) to generate "reasonable" initial inequalities. This selection helped to increase the lower bound significantly in the early iterations and to speed up the overall running time.

Initial variables. For large combinatorial optimization problems the number of variables of the LP-relaxation may be tremendous. A helpful trick is to restrict the LPs to "promising" variables that are chosen heuristically. Of course in the end, this planned error has to be repaired. We will show later how this is done. For the travelling salesman problem, for instance, a typical choice are the 2 to 10 nearest neighbors of any node and the edges of several heuristically generated tours. For a 3000 city TSP instance, the number of variables of the initial LP can be restricted from about 4.5 million to less than 10 thousand this way; see Applegate et al. (1994), Grötschel and Holland (1991), and Padberg and Rinaldi (1991) for descriptions of variable reduction strategies for the TSP.

There are further preprocessing techniques that, depending on the type of problem and the special structure of the instances, can be applied. These techniques are vital in many cases to achieve satisfactory running times in practice. Particularly important are techniques for structurally reducing instance sizes by decomposition, for detecting logical dependencies, implicit relations, and bounds that can be used to eliminate variables or forget certain constraints forever. For space reasons we are unable to outline all this here.

Cutting plane generation. The core of a cutting plane procedure is of course the identification of violated inequalities. Assume that we have made our choice of initial constraints and have solved the initial linear program $\max\{c^{\mathrm{T}}x \mid A_{\mathrm{I}}x \leq b_{\mathrm{I}}\}$. In further iterations we may have added additional constraints so that the current linear program has the form

$$\begin{aligned} &\text{maximize} && c^{\mathrm{T}}x \\ &\text{subject to} && Ax \leq b\,. \end{aligned}$$

We solve this LP and suppose that y is an optimum solution. If y is the incidence vector of a feasible solution of our combinatorial problem we are done. Otherwise we want to check whether y satisfies all the constraints in $A_2x \leq b_2, \ldots, A_kx \leq b_k$. We may check certain small classes by substituting y into all inequalities. But, in general, we will run all the separation routines (exact and heuristic) that we have, to find as many inequalities violated by y as possible. It is a very good idea to use several different heuristics even for classes of inequalities for which exact separation algorithms are available. The reason is that exact routines typically find only a few violated constraints (the most violated ones), while separation heuristics often come up with many more and differently structured constraints.

To keep the linear programs small one does also remove constraints, for instance those that are not tight at the present solution. It is sometimes helpful to keep these in a "pool" since an optimum solution of a later iteration might violate it again, and scanning the pool might be computationally cheaper than running elaborate separation routines (see Padberg and Rinaldi 1991).

In the initial phase of a cutting plane procedure the separation routines may actually produce thousands of violated constraints. It is then necessary to select "good ones" heuristically, again, to keep the LPs at a manageable size, see Grötschel et al. (1984) for this issue.

Another interesting issue is the order in which exact separation routines and heuristics are called. Although that may not seem to be important, running time factors of 10 or more may be saved by choosing a suitable order for these and strategies to give up calling certain separation heuristics. An account of this matter is given in Barahona et al. (1988).

There are more aspects that have to be considered, but we are unable to cover all these topics here. It is important to note that, at the present state of the art, there are still no clear rules as to which of these issues are important or almost irrelevant for a combinatorial optimization problem and its LP relaxation considered. Many computational experiments with data from practical instances of realistic sizes are necessary to obtain the right combination of methods and "tricks".

Pricing variables. In the cutting plane procedure we have now iteratively called the cutting plane generation methods, added violated inequalities, dropped a few constraints and repeated this process until we either found an optimum integral solution or stopped with an optimum fractional solution y for which no violated constraint could be found by our separation routines. Now we have to consider the "forgotten variables". This is easy. For every initially discarded variable we generate the column corresponding to the present linear constraint system and compute its reduced costs by standard LP techniques. If all reduced costs come out with the correct sign we have shown that the present solution is also optimum for the system consisting of all variables. If this is not the case we add all (or some, if there are too many) of the variables with the wrong sign to our current LP and repeat the cutting plane procedure. In fact, using reduced cost criteria one can also show that some variables can be dropped because they can provably never appear in any optimum solution or that some variables can be fixed to a certain value.

Branch-and-cut. This process of iteratively adding and dropping constraints and variables may have to be repeated several times before an optimum solution y of the full LP is found. However, for large instances this technique is by far superior to the straightforward method of considering everything at once. If the final solution y is integral, our combinatorial optimization problem is solved. If it is not integral, we have to resort to branch-and-bound, see section 4. There are various ways to mix cutting plane generation with branching, to use fractional LP-solutions for generating integral solutions heuristically, etc. It has thus become popular to call the whole approach described here *branch-and-cut*.

Clearly, this tremendous theoretical and implementational effort only pays if the bounds for the optimum solution value obtained this way are very good. Computational experience has shown that, in many cases, this is indeed so. We refer the interested readers to more in-depth surveys on this topic such as Grötschel and Padberg (1985), Padberg and Grötschel (1985), and Jünger et al. (1995) for the TSP, or to papers describing the design and implementation of a cutting plane algorithm for a certain practically relevant, hard combinatorial optimization problem. These papers treat many of the issues and little details we were unable to cover here. Among these papers are: Applegate et al. (1994), Grötschel and Holland (1991), Padberg and Rinaldi (1991) for the TSP (the most spectacular success of the cutting plane technique has certainly been achieved here); Barahona et al. (1988) for the max-cut problem with applications to ground states in spin glasses and via minimization in VLSI design; Grötschel et al. (1984) for the linear ordering problem with applications to triangulation of input–output matrices and ranking in sports; Hoffman and Padberg (1993) for the set partitioning problem with applications to airline crew scheduling; Grötschel and Wakabayashi (1989) for the clique partitioning problem with applications to clustering in biology and the social sciences; Grötschel et al. (1992b) for certain connectivity problems with applications to the design of survivable telecommunication networks; Grötschel et al. (1992a) for the Steiner tree packing problem with applications to routing in VLSI design.

The LP-solvers used in most of these cases are advanced implementations of the simplex algorithm such as Bixby's CPLEX or IBM's OSL. Investigations of the use of interior point methods in such a framework are on the way. A number of important issues like addition of rows and columns and postoptimality analysis, warm starts, etc. are not satisfactorily solved yet. But combinations of the two approaches may yield the LP-solver of the future for this approach.

Linear programming in heuristics. So far, we have used linear programming as a dual heuristic, to obtain upper bounds on (say) the maximum value of an integer program. But solving the linear relaxation of an optimization problem also provides primal information, in the sense that it can be used to obtain a (primal) heuristic solution.

Of course, solving the linear relaxation of an integer linear program, we may be lucky and get an integer solution right away. Even if this does not happen, we may find that some of the variables are integral in the optimum solution of the linear relaxation, and we may try to fix these variables at these integral values. This is in general not justified; a notable special case when this can be done was found by Nemhauser and Trotter (1974), who proved the following. Consider a graph $G = (V, E)$ and the usual integer linear programming formulation of the stable set problem:

$$\begin{aligned} \text{maximize} \quad & \sum_{i \in V} x_i \\ \text{subject to} \quad & x_i \geqslant 0 \; (i \in V) \\ & x_i + x_j \leqslant 1 \; (ij \in E) \\ & x \text{ integral}. \end{aligned} \tag{9}$$

Let x^* be an optimum solution of the linear relaxation of this problem. Then there exists an optimum solution x^{**} of the integer program such that $x_i^* = x_i^{**}$ for all i for which x_i^* is an integer.

In general, we can obtain a heuristic primal solution by fixing those variables that are integral in the optimum solution of the linear relaxation, and rounding the remaining variables "appropriately". Properties of this heuristic were studied in detail by Raghavan and Thompson (1987). However, it seems that this natural and widely used scheme for a heuristic is not sufficiently analyzed. Here we discuss some results where linear programming combined with appropriate rounding procedures gives a provably good primal heuristic.

A polynomial approximation scheme for bin packing. The following polynomial time bin packing heuristic, due to Fernandez de la Vega and Lueker (1981), has asymptotic performance ratio $1 + \varepsilon$, where $\varepsilon > 0$ is any fixed number. A more refined application of this idea gives a heuristic that packs the weights into $k_{\text{opt}} + O(\log^2(k_{\text{opt}}))$ bins (Karmarkar and Karp 1982).

First, we solve the following restricted version. We are given integers $k, m > 0$, weights $1/k < a_1 < \cdots < a_m \leqslant 1$ and a multiplicity n_j for each weight a_j. Let k_{opt}

be the minimum number of bins into which n_j copies of a_j $(j = 1, \ldots, m)$ can be packed. Then we can pack the weights into $k_{\text{opt}} + m$ bins, in time polynomial in n and $\binom{m+k}{k}$.

To obtain such a packing, let us first generate all possible combinations (with repetition) of the given weights that fit into a single bin. Since each weight is at least $1/k$, such a combination has at most k elements, and hence the number of different combinations is at most $\binom{m+k}{k}$, and they can be found by brute force. Let $T_1, \ldots, T_N$ be these combinations. Each T_i can be described by an integer vector $(t_1^i, t_2^i, \ldots, t_m^i)$, where t_j^i is the number of times weight a_j appears in combination i.

Consider the following linear program:

$$\begin{aligned} &\text{minimize} && \sum_{i=1}^{N} y_i \\ &\text{subject to} && y_i \geqslant 0 \\ & && \sum_{i=1}^{N} t_j^i y_i \geqslant n_j \ (j = 1, \ldots, m)\,. \end{aligned} \tag{10}$$

Let Y denote the optimum value. Every packing of the given weights into bins gives rise to an (integral) solution of this linear program (y_i is the number of times combination T_i is used), hence

$$k_{\text{opt}} \geqslant \mathrm{Y}\,.$$

On the other hand, let y^* be an optimum basic solution of (10), and consider $\lceil y_i^* \rceil$ bins packed with combination T_i. Since at most m of the y_i^* are non-zero, we get a total of $\sum_i \lceil y_i^* \rceil < Y + m \leqslant k_{\text{opt}} + m$ bins, which clearly accommodate the whole list.

Now let $0 < x_1 \leqslant x_2 \leqslant \cdots \leqslant x_n \leqslant 1$ be an arbitrary list L of weights, and let $0 < \varepsilon < 1$ be also given. Set $w := \sum_i x_i$, and define l by $x_l < \varepsilon/2 \leqslant x_{l+1}$ (set $l := 0$ if $x_1 \geqslant \varepsilon/2$). Set $a_i := x_{l+ih}$ $(i = 1, \ldots, m)$, where $h := \lceil \varepsilon w \rceil$ and $m := \lfloor (n-l)/h \rfloor$. Consider a list L' consisting of h copies of each a_i and $n - l - hm$ copies of 1. Let k'_{opt} be the minimum number of bins into which L' can be packed; by the solution of the restricted problem described above, we can pack L' into $k'_{\text{opt}} + m$ bins in polynomial time. Trivially, we get from this a packing of the weights $x_{l+1}, \ldots, x_n$ into $k'_{\text{opt}} + m$ bins. The remaining (small) weights $x_1, \ldots, x_l$ are packed by FIRST-FIT into the slacks of these bins and, if necessary, into new bins.

To compare the number k_{heur} of bins used this way with k_{opt}, we distinguish two cases. If, in the last step of FIRST-FIT, we had to open a new bin, then every bin (except possibly the last one) is filled up to $1 - \varepsilon/2$, and hence

$$k_{\text{heur}} \leqslant 1 + \frac{1}{1 - \varepsilon/2}\, w\,.$$

Since w is a trivial lower bound on k_{opt}, this shows that

$$k_{\text{heur}} \leqslant (1 + \varepsilon) k_{\text{opt}} + 1\,.$$

So assume that we do not open a new bin in the last phase, and hence we use at most $k'_{\rm opt}+m$ bins. To compare this with $k_{\rm opt}$, consider an optimum packing of L. Then from the list L', the h copies of a_1 can be put in the place of $x_{l+h+1},\ldots,x_{l+2h}$, the h copies of a_2 can be put in the place of $x_{l+2h+1},\ldots,x_{l+3h}$ etc. At the end we are left with h weights, which we can accommodate using h new bins. Hence

$$k'_{\rm opt}\leq k_{\rm opt}+h\,,$$

and so

$$k_{\rm heur}\leq k'_{\rm opt}+m\leq k_{\rm opt}+h+m\,.$$

Since

$$h\leq \varepsilon w+1\leq \varepsilon k_{\rm opt}+1\,,$$

and

$$m\leq \frac{n-l}{h}<\frac{w}{\varepsilon/2}\Big/(\varepsilon w)=\frac{2}{\varepsilon^2}=\mathrm{O}(1)\,,$$

this proves that the asymptotic performance ratio is at most $1+\varepsilon$ as claimed.

Blocking sets in hypergraphs with small Vapnik–Červonenkis dimension. The following discussion is based on ideas of Vapnik and Červonenkis, which have become very essential in a number of areas in mathematics (statistics, learning theory, computational geometry). Here we use these ideas to design a randomized heuristic for finding a blocking set in a hypergraph.

Let $(V,\mathscr{H})$ be a hypergraph and consider an optimum fractional blocking set of $\mathscr{H}$, i.e., an optimum solution x^* of the linear program

$$\begin{aligned}\text{minimize}\quad & \sum_{i\in V} x_i\\ \text{subject to}\quad & x_i\geq 0\ (i\in V)\\ & \sum_{i\in E} x_i\geq 1\ (E\in\mathscr{H})\,.\end{aligned}\tag{11}$$

The optimum value of program (11) is the *fractional blocking number* $\tau^*:=\tau^*(\mathscr{H})$, which can serve as a lower bound on the *covering* (or blocking) *number* $\tau(\mathscr{H})$. Now we use this linear program to obtain a heuristic solution.

Consider an optimum solution x, and define $p_i=x_i/\tau^*$. Then $(p_i\colon i\in V)$ can be viewed as a probability distribution on V, in which every edge $E\in\mathscr{H}$ has probability at least $1/\tau^*$. Let us generate nodes $v_1,v_2,\ldots$ independently from this distribution, and stop when all edges are covered. It is easy to see that with very large probability we stop in a polynomial number of steps, so this procedure is indeed a (randomized) blocking set heuristic, which we call the random node heuristic. Let $t_{\rm heur}$ be the size of the blocking set produced (this is a random variable). What is the expected performance ratio of the random node heuristic, i.e., the ratio $\mathrm{E}(t_{\mathrm{heur}})/\tau(\mathcal{H})$?

It is clear that if we consider a particular edge E, then it will be hit by one of the first $k\tau^*$ nodes with probability about $1 - 1/\mathrm{e}^k$. Hence if $k > \ln|\mathcal{H}|$, then the probability that every edge is hit is more than $\frac{1}{2}$. Hence we get the inequality

$$t_{heur} \leq (\ln|\mathcal{H}|)\tau^* \leq (\ln|\mathcal{H}|)\tau ,$$

i.e., we obtain the bound $\ln|\mathcal{H}|$ for the performance ratio of the random node heuristic. This is not interesting, however, since the greedy heuristic does better (see Theorem 2.2).

Adopting a result of Haussler and Welzl (1987) from computational geometry (which in turns is an adaptation of the work of Vapnik and Červonenkis in statistics, see Vapnik 1982), we get a better analysis of the procedure. The *Vapnik–Červonenkis dimension* of a hypergraph $(V, \mathcal{H})$ is the size of the largest set $S \subseteq V$ such that for every $T \subseteq S$ there is an $E \in \mathcal{H}$ such that $T = S \cap E$.

Theorem 8.3. *Let $\mathcal{H}$ be a hypergraph with Vapnik–Červonenkis dimension d and fractional blocking number τ^*. The expected size of the blocking set returned by the random node heuristic is at most $16d\tau^* \log(d\tau^*)$.*

Proof. We prove that if we select $N := \lceil 8d\tau^* \log(d\tau^*)\rceil$ nodes from the distribution p, then with probability more than $\frac{1}{2}$, every edge of $\mathcal{H}$ is met. Hence it follows easily that $\mathrm{E}(t_{\mathrm{heur}}) \leq 2N$.

The proof is not long but tricky. Choose N further nodes $v_{N+1}, \ldots, v_{2N}$ (independently, from the same distribution). Set $s = N/(2\tau^*)$. Assume that there exists a set $E \in \mathcal{H}$ such that $E \cap \{v_1, \ldots, v_N\} = \emptyset$. Chebychev's inequality gives that for any such edge $E \in \mathcal{H}$,

$$\mathrm{Prob}(|E \cap \{v_{N+1}, \ldots, v_{2N}\}| \geq s) > \tfrac{1}{2} .$$

Hence we obtain

$$\begin{aligned}&\mathrm{Prob}(\exists E\colon E \cap \{v_1, \ldots, v_N\} = \emptyset, |E \cap \{v_{N+1}, \ldots, v_{2N}\}| > s)\\ &\quad > \tfrac{1}{2}\,\mathrm{Prob}(\exists E\colon E \cap \{v_1, \ldots, v_N\} = \emptyset) .\end{aligned}$$

We estimate the probability on the left-hand side from above as follows. We can generate a $(2N)$-tuple from the same distribution if we first generate a set S of $2N$ nodes as before, and then randomly permute them. For a given $E \in \mathcal{H}$ that meets S in at least s elements, the probability that after this permutation E avoids the first half of S is at most

$$\frac{\binom{2N-s}{N}}{\binom{2N}{N}} \leq \left(1 - \frac{s}{2N}\right)^N < \mathrm{e}^{-s/2} .$$

We do not have to add up this bound for all E, only for all different intersections

$E \cap S$. The number of sets of the form $E \cap S$ is at most

$$\binom{2N}{d} + \binom{2N}{d-1} + \cdots + 1 < (2N)^d ,$$

by the Sauer-Shelah theorem (see chapter 24, section 4). Hence the probability that there is an edge $E \in \mathscr{H}$ that meets S in at least s elements but avoids the first half is at most $(2N)^d \mathrm{e}^{-s/2} < 1/4$.

Hence

$$\mathrm{Prob}(\exists E : E \cap \{v_1, \ldots, v_N\} = \emptyset) < \tfrac{1}{2} . \qquad \square$$

With some care, the upper bound can be improved to (essentially) $\mathrm{O}(d\tau^* \log \tau^*)$, which is best possible in terms of these parameters, see Komlós et al. (1992).

Approximating a cost-minimal schedule of parallel machines. Machine scheduling problems arise in hundreds of versions and are a particular "playground" for approximation techniques. We outline here an LP-based heuristic for the following problem of scheduling parallel machines with costs (this problem is also called the *generalized assignment problem*). Suppose that we have a set J of n independent jobs, a set M of m unrelated machines, and we want to assign each job to one of the machines. Assigning job j to machine i has a certain cost c_{ij} and takes a certain time p_{ij}. Our task is to find an assignment of jobs to machines such that no machine gets more load than a total of T time, and the total cost does not exceed a given bound C, i.e., we look for a job assignment with maximum time load (*makespan*) at most T and cost at most C.

If all the p_{ij} are the same, then this is a weighted bipartite matching problem, and so can be solved in polynomial time. However, for general p_{ij}, the problem is NP-hard. Since there are two parameters (T and C), there are several ways to formulate what an approximate solution means, and there are various algorithms known to find them. Each of these is based on solving a linear relaxation of the problem and then "rounding" the solution appropriately; this technique was introduced by Lenstra et al. (1990). A combinatorially very interesting "rounding" of the solution of the linear relaxation was used by Shmoys and Tardos (1993), which we now sketch.

Consider the following linear program:

$$\begin{aligned}
&\min c^{\mathrm{T}} x := \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} , \\
&\sum_{j=1}^{m} p_{ij} x_{ij} \leq T , && \text{for } i = 1, \ldots, m , \\
&\sum_{i=1}^{m} x_{ij} = 1 , && \text{for } j = 1, \ldots, n , \\
&x_{ij} \geq 0 , && \text{for } i = 1, \ldots, m,\ j = 1, \ldots, n ,
\end{aligned}$$

$$x_{ij} = 0\,, \qquad \text{if } p_{ij} > T, i = 1, \ldots, m, j = 1, \ldots, n\,. \tag{12}$$

Clearly, every integral solution y of (12) with cost $c^{\mathrm{T}}y \leqslant C$ provides a feasible solution of the generalized assignment problem, and thus, (12) is a natural LP-relaxation of the generalized assignment problem. (The explicit inclusion of the last condition plays an important role in the approximation algorithm.) Let us replace the right hand side T of the first m inequalities of (12) by $2T$ and let us denote this new LP by (12′). Now Shmoys and Tardos prove the following. *If* (12) *has a (possibly fractional) solution x^* with cost $c^* := c^{\mathrm{T}}x^*$ then* (12′) *has a* 0/1-*solution with cost c^**. In other words, if the LP (12) has a solution with cost at most C, then there is an assignment of jobs to machines with the same cost (at most C) and makespan at most $2T$.

The trick is, of course, in "rounding" the real solution x^*. This is done by using x^* to construct an auxiliary bipartite graph and then finding a minimum cost matching in this graph (cf. section 9), which then translates back to an assignment of cost at most C and makespan at most $2T$. These details must be omitted here, but note that the "rounding" involves a nontrivial graph-theoretic algorithm.

9. Changing the objective function

Consider an optimization problem in which the objective function involves some "weights". One expects that if we change the weights "a little", the optimum solutions do not change, or at least do not change "too much". It is surprising how far this simple idea takes us: it leads to efficient algorithms, motivates linear programming, and is the basis of fundamental general techniques (scaling, Lagrangean relaxation, strong polynomiality).

Kruskal's algorithm revisited. Let $G = (V, E)$ be a connected graph and $c: V \rightarrow \mathbb{Z}$, the length function of its edges. We want to find a shortest spanning tree. Clearly, adding a constant to all edges does not change the problem in the sense that the set of optimum solutions remains the same. Thus, we may assume that the lengths are non-negative.

Now let us push this idea just a bit further: we may assume that the shortest edge has length 0. Then it is easy to see that shrinking this edge to a single node does not alter the length of the shortest spanning tree. We can shift the edge-weights again so that the minimum length of the remaining edges is 0; hence, we may contract another edge, etc.

It is easy to see that this algorithm to construct a shortest spanning tree is actually Kruskal's algorithm in disguise: the first edge contracted is the shortest edge; the second, the shortest edge not parallel to the first, etc. (Or is greediness a disguise of this argument?)

Minimum weight perfect matching in a bipartite graph. Given a complete bipartite graph $G=(V,E)$ with bipartition (U,W), where $|U|=|W|$, and a cost function $w: E\to\mathbb{Z}$, we want to find a perfect matching M of minimum weight.

The idea is similar to the one mentioned in the previous section. By adding the same constant to each weight, we assume that all the weights are non-negative. But now we have more freedom: if we add the same constant to the weights of all edges incident with a given node, then the weight of every perfect matching is also shifted by the same constant, and so the set of optimum solutions does not change. Our aim is to use this transformation until a perfect matching with total weight 0 is obtained; this is then triviually optimal.

Let G_0 denote the graph formed by edges with weight 0. Using the unweighted bipartite matching algorithm, we can test whether G_0 has a perfect matching. If this is the case, we are done; else, the algorithm returns a set $X\subseteq U$ such that the set of neighbors $N_{G_0}(X)$ of X is smaller than X. If ε is the minimum weight of any edge from X to $W\backslash N_{G_0}$, we add ε to all the edges out of $N_{G_0}(X)$ and $-\varepsilon$ to all the edges out of X. This transformation preserves the values of the edges between X and $N_{G_0}(X)$, and creates at least one new node connected to X by a 0-edge. Any perfect matching changes its weight by the same value $-\varepsilon|X|+\varepsilon|N_{G_0}(X)|<0$.

It remains to show that this procedure terminates, and to estimate the number of iterations it takes. One way to show this is to remark that at each iteration, either the maximum size of an all-0 matching increases, or it remains the same, but then the set X returned by the unweighted bipartite matching algorithm (as described in chapter 3) increases. This gives an $O(n^2)$ bound on the number of iterations.

One can read off from this algorithm Egerváry's min–max theorem on weighted bipartite matchings (see chapter 3 for extensions of the algorithm and the theorem to non-bipartite graphs).

Theorem 9.1. *If $G=(V,E)$ is a bipartite graph with bipartition (U,W), where $|U|=|W|$, and $w: E\to\mathbb{Z}$ is a cost function, then the minimum weight of a perfect matching is the maximum of $\sum_{i\in V}\pi_i$,* where *$\pi: V\to\mathbb{Z}$ is a weighting of the nodes such that $\pi_i+\pi_j\leqslant c_{ij}$ for all $ij\in E$.*

Optimum arborescences. Given a digraph $G=(V,A)$ and a root $r\in V$, an *arborescence* is a spanning tree whose edges are oriented away from r. Let us assume that G contains an arborescence with root r. (This is easily checked.) Let $l: A\to\mathbb{Z}$ be an assignment of "lengths" to the arcs. The *optimum arborescence problem* is to find an arborescence of minimum length (see also chapter 30). An optimum arborescence can be found efficiently by the following algorithm due to Edmonds (1967a).

Again, we may assume that the lengths are non-negative, since this can be achieved by adding a constant to every arc length.

Consider all the arcs of G going into some node $v\neq r$. Any arborescence will contain exactly one of these arcs. Hence we may add a constant to the length of all the arcs going into v without changing the set of optimum arborescences.

For every $v\neq r$, we add a constant to the arcs going into v so that the minimum

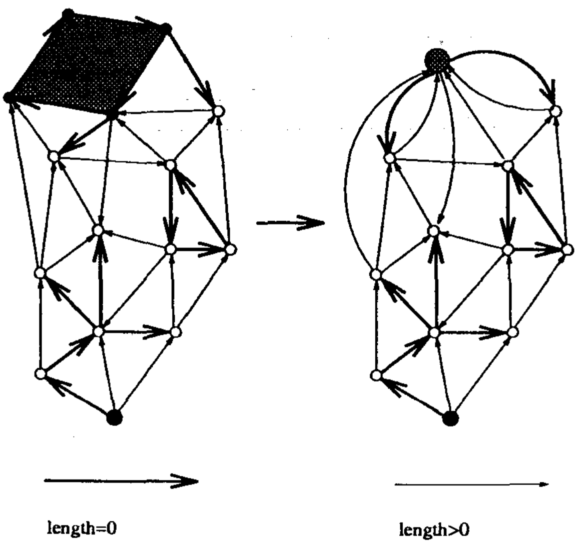


Figure 9.1. The optimum arborescence algorithm.

length of arcs entering any given vertex is 0. Consider the subgraph of all the arcs of length 0. If there exists an arborescence contained in this subgraph, we are done. Otherwise, there must be a cycle of 0-arcs (fig. 9.1).

We contract this 0-cycle, to get the graph G'. It is easy to check that the minimum length of an arborescence in G' is equal to the minimum length of an arborescence in G. Thus we can reduce the problem to a smaller problem, and then proceed recursively. One reduction requires $\mathrm{O}(m)$ time, so this leads to an $\mathrm{O}(mn)$ algorithm.

Similarly as in the case of the weighted bipartite matching algorithm, we can use this algorithm to derive Fulkerson's optimum arborescence theorem (chapter 2, Theorem 6.4).

More involved applications of the idea of shifting the objective function without changing the optimum solutions include the out-of-kilter method (see chapter 2, section 5).

Scaling I: From pseudopolynomial to polynomial. Consider a finite set E and a collection $\mathcal{F}$ of subsets of E. (in the cases of interest here, $\mathcal{F}$ is implicitly given and may have size exponentially large in $|E|$, e.g., the set of all Eulerian subgraphs of a digraph). Let a weight function $w: E \to \mathbb{Q}$ be also given. Our task is to find a member $X \in \mathcal{F}$ with maximum weight $w(X) := \sum_{e \in X} w(e)$.

Let us round each weight $w(e)$ $(e \in E)$ to the nearest integer. Does this change the set of optimum solutions? On course, it may; but in several situations, connections between the original and the rounded problem can be established so that solving the rounded (and, sometimes, simpler) problem helps in the solution of the original. Before rounding, we may of course multiply each $w(e)$ by the same positive scalar; combined with rounding, this becomes a powerful technique. It was introduced by Edmonds and Karp (1972) to show the polynomial time solvability of the minimum cost flow problem. Since then, scaling has become one of the most fundamental tools in combinatorial optimization, in particular in flow theory (see, e.g., Goldberg et al. 1990).

We illustrate the method on a simple, yet quite general example.

Theorem 9.2. *Let Ψ be a family of hypergraphs and consider the optimization problem* $\max\{w(X)\colon X \in \mathcal{F}\}$ *for members $(E, \mathcal{F}) \in \Psi$ and objective functions $w\colon E \to \mathbb{Z}_+$. Assume that there exists an "augmentation", i.e., an algorithm that checks whether $X \in \mathcal{F}$ is optimal and if not, returns an $X' \in \mathcal{F}$ with $w(X') > w(X)$; also assume that the augmentation algorithm runs in time polynomial in $\langle w \rangle$. Then the optimization problem can be solved in time polynomial in $\langle w \rangle$.*

Proof. Note that a pseudopolynomial algorithm for this problem is obvious: start with any $X \in \mathcal{F}$ and augment until optimality is achieved. The number of augmentations is trivially bounded by $w(E)$. (Another obvious bound is $2^{|E|}$.) It is easy to construct examples where this trivial algorithm is not polynomial.

To achieve this in polynomial time, let $k := \max_e \lfloor \log w(e) \rfloor$, and define new objective functions $w_j := \lfloor w/2^{k-j} \rfloor$. We solve the optimization problem for the objective function w_0, then for $w_1, \ldots,$ finally for $w_k = w$. Since w_0 is 0/1-valued, we can apply the pseudopolynomial algorithm to find the optimizing set.

Assume that we have an optimizing set X_j for w_j. This is of course also optimal for $2w_j$, which is very close to w_{j+1}: we have, for each $e \in E$,

$$2w_j(e) \leq w_{j+1}(e) \leq 2w_j(e) + 1 .$$

Hence, we have for any set $X \in \mathcal{F}$,

$$w_{j+1}(X) \leq 2w_j(X) + n \leq 2w_j(X_j) + n \leq w_{j+1}(X_j) + n .$$

Thus, the set X_j is almost optimal for the objective function w_{j+1}, and the trivial algorithm starting with X_j will maximize w_{j+1} in at most n iterations. Therefore, w will be maximized in a total of $O(nk)$ iterations. □

As an example, consider the problem of finding an Eulerian subdigraph of maximum weight in a directed graph $D = (V, A)$ with arc weights w_a. Then $(A, \mathcal{F})$, where $\mathcal{F} := \{C \subseteq A\colon C \text{ Eulerian}\}$, is a hypergraph. To apply Theorem 9.2 we have to design an augmentation subroutine. Given some Eulerian subdigraph C we construct an auxiliary digraph D_C by reversing the arcs in $A \setminus C$ and changing the signs of the weights on these edges. C is not a maximum weight Eulerian digraph if and only if D_C contains a directed circuit of negative total weight. Such a circuit can be found in polynomial time by shortest path techniques.

For more involved applications of these scaling techniques see chapter 2, section 5.

Scaling II: From polynomial to strongly polynomial. Strong polynomial solvability of a problem is often much more difficult to prove than polynomial solvability; for example, it is not known whether linear programs can be solved in strongly polynomial time. It is therefore remarkable that Frank and Tardos (1987) showed that, for a large class of combinatorial optimization problems, polynomial solvability implies strong polynomial solvability (see also chapter 30).

Theorem 9.3. *Let Ψ be a family of hypergraphs and assume that there exists an algorithm to find* $\max\{\sum_{e\in X} w(e) : X \in \mathcal{F}\}$ *for every* $(E, \mathcal{F}) \in \Psi$ *and* $w: E \to \mathbb{Z}$ *in time polynomial in* $\langle w \rangle$. *Then there exists a strongly polynomial algorithm for this maximization problem.*

Proof. The goal is to find an algorithm in which the number of arithmetic operations is bounded by a polynomial in $n = |E|$, and does not depend on $\langle w \rangle$ (we also need that the numbers involved do not grow too wild, but this is easy to check). So the bad case is when the entries of w are very large. Frank and Tardos give an algorithm that replaces w by an integer vector w' such that every entry of w' has at most $O(n^3)$ digits, and w and w' are maximized by the same members of $\mathcal{F}$.

The key step is the construction of the following *diophantine expansion* of the vector w:

$$w = \lambda_1 u_1 + \lambda_2 u_2 + \cdots + \lambda_n u_n ,$$

where $u_1, \ldots, u_n$ are integral vectors with $1 \leq \|u_i\|_\infty \leq 4^{n^2}$, and the coefficients λ_i are rational numbers that decrease very fast:

$$|\lambda_{i+1}| < \frac{1}{2n\|u_{i+1}\|_\infty} |\lambda_i| \quad (i = 1, \ldots, n-1) .$$

Such an expansion can be constructed using the simultaneous diophantine approximation algorithm (see chapter 19). This expansion has the property that for any two sets $X, Y \subseteq E$,

$$w(X) \leq w(Y) \iff u_i(X) \leq u_i(Y) \text{ for } i = 1, \ldots, n .$$

Now letting

$$w' := 8^{n^3} u_1 + 8^{n^3 - n^2} u_2 + 8^{n^3 - 2n^2} u_3 + \cdots + 8^{n^2} u_n ,$$

we have for any two sets $X, Y \subseteq E$,

$$w(X) \leq w(Y) \iff w'(X) \leq w'(Y) .$$

Thus w and w' are optimized by the same sets, and we can apply our polynomial time algorithm to maximize w'. Since $\langle w' \rangle = O(n^3)$, this algorithm will be strongly polynomial. □

Applying the result to our previous example, we obtain a strongly polynomial algorithm for the maximum weight Eulerian subdigraph problem. More generally, this technique yields strongly polynomial algorithms, among others, for linear programs with $\{-1, 0, 1\}$-matrices (e.g., for the minimum cost flow problem).

Poljak (1993) applied an even more general version of scaling to show that the single exchange heuristic of the max-cut problem is strongly polynomial for cubic graphs (while exponential for 4-regular graphs). Let $G = (V, E)$ be a graph and $c: E \rightarrow \mathbb{Z}$, a weighting of its edges. The idea is that the run of the heuristic is determined if we know, for each node v and each partition of the edges incident with v into two classes, which class has larger weight. So we consider a family of inequalities, each of which is of the type

$$x_i + x_j \geq x_k \quad \text{or} \quad x_i + x_j + 1 \leq x_k \tag{13}$$

(where i, j and k are three edges adjacent to a node). We know that this system has a solution (the original weights). Poljak proves that then the system has an integral solution with $1 \leq |x_i| \leq 2|V| - 1$ for all i. Replacing the original weights with these new weights, the single exchange heuristic runs as before, but now it clearly terminates in $O(|V|^2)$ time.

Scaling III: Heuristics. Recall from section 5 that the 0/1-knapsack problem is NP-hard, but it is polynomially solvable if the weight coefficients a_i are given in unary notation. This fact was combined with a scaling technique by Ibarra and Kim (1975) to design a fully polynomial approximation scheme for the knapsack problem.

Fix any $\varepsilon > 0$. We may assume that $c_1 \geq c_2 \geq \cdots \geq c_n$. Let C_{opt} denote the optimum value of the knapsack, and let C be an upper bound on C_{opt}; a good value for C can, e.g., be found by running the greedy heuristic for the knapsack problem, for which Theorem 2.7 gives

$$\frac{C}{2} \leq C_{\mathrm{opt}} \leq C .$$

Let $0 \leq m \leq n$ be the largest index such that $c_m > \varepsilon C/4$, and define

$$\bar{c}_i = \left\lfloor \frac{8c_i}{\varepsilon^2 C} \right\rfloor .$$

Clearly $\bar{c}_i \leq 8/\varepsilon^2$, and so these numbers are polynomially bounded in $1/\varepsilon$.

The idea is to solve a knapsack problem omitting the "small" weights and replacing the "big" weights $c_1, \ldots, c_m$ by their unary approximations $\bar{c}_1, \ldots, \bar{c}_m$. Then we use the "small" weights to fill up greedily as much of the remaining space as possible.

For every integral value d with $0 \leq d \leq 8/\varepsilon^2$, we determine a solution x^d of the knapsack problem, by solving the following auxiliary optimization problem:

$$\begin{aligned}
\text{minimize} \quad & \sum_{i=1}^{m} a_i x_i \\
\text{subject to} \quad & \sum_{i=1}^{m} \bar{c}_i x_i = d\,, \\
& x_i \in \{0,1\}, i = 1, \ldots, m\,. \qquad (14)
\end{aligned}$$

This is basically a subset-sum problem with a linear objective function and can be solved, by the same dynamic programming argument, in polynomial time. (In fact, we get the optimum solution for all values of d in a single run, which takes $O(n/\varepsilon^2)$ time for the execution of all problems (14).) Let $x_1^d, \ldots, x_m^d$ be the solution found (if any exists).

Now we choose the remaining variables. These variables $x_{m+1}, \ldots, x_n$ must satisfy the following constraints:

$$\begin{aligned}
& \sum_{i=m+1}^{n} a_i x_i \leqslant b - \sum_{i=1}^{m} a_i x_i^d\,, \\
& x_i \in \{0,1\}\,, \qquad (15)
\end{aligned}$$

and we want to maximize

$$\sum_{i=m+1}^{n} c_i x_i\,.$$

If the right-hand side in (15) is non-negative, then this is just another (auxiliary) knapsack problem, which we solve by the greedy algorithm in $O(n)$ time for each d (thus using $O(n/\varepsilon^2)$ time in total). Let $x_{m+1}^d, \ldots, x_n^d$ be the solution of the knapsack obtained (if it exists, i.e., if (15) has non-negative right-hand side).

Theorem 9.4. *For at least one d with $0 \leqslant d \leqslant 8/\varepsilon^2$, the solution $(x_1^d, \ldots, x_n^d)$ exists, and*

$$\sum_{i=1}^{n} c_i x_i^d \geqslant (1-\varepsilon) C_{\mathrm{opt}}\,.$$

Proof. Let $y_1, \ldots, y_n$ be a (true) optimum solution of the original knapsack problem, and consider the value

$$d := \sum_{i=1}^{m} \bar{c}_i y_i\,.$$

Clearly

$$d \leqslant \frac{8}{\varepsilon^2 C} \sum_{i=1}^{m} c_i y_i \leqslant \frac{8}{\varepsilon^2}\,.$$

Fix this choice of d. Then trivially $x_1^d, \ldots, x_m^d$ exists, and by their optimality for the auxiliary subset-sum problem (14), we have

$$\sum_{i=1}^{m} a_i x_i^d \leqslant \sum_{i=1}^{m} a_i y_i \leqslant \sum_{i=1}^{n} a_i y_i \leqslant b\ .$$

Thus for this d, (15) has non-negative right-hand side and the solution $(x_1^d, \ldots, x_n^d)$ exists. Moreover,

$$\begin{aligned}\sum_{i=1}^{m} c_i x_i^d &\geqslant \frac{\varepsilon^2 C}{8} \sum_{i=1}^{m} \bar{c}_i x_i^d = \frac{\varepsilon^2 C}{8} \sum_{i=1}^{m} \bar{c}_i y_i \\ &\geqslant \sum_{i=1}^{m} c_i y_i - \frac{\varepsilon^2 C}{8} \sum_{i=1}^{m} y_i\ .\end{aligned} \tag{16}$$

Here

$$\sum_{i=1}^{m} y_i \leqslant \frac{1}{c_m} \sum_{i=1}^{m} c_i y_i \leqslant \frac{4}{\varepsilon C} C_{\text{opt}}\ ,$$

and so

$$\sum_{i=1}^{m} c_i x_i^d \geqslant \sum_{i=1}^{m} c_i y_i - \frac{\varepsilon}{2} C_{\text{opt}}\ .$$

Furthermore, observe that $(y_{m+1}, \ldots, y_n)$ is a solution of (15), and hence by Theorem 2.7,

$$\sum_{i=m+1}^{n} c_i x_i^d \geqslant \sum_{i=m+1}^{n} c_i y_i - c_{m+1} \geqslant \sum_{i=m+1}^{n} c_i y_i - \frac{\varepsilon}{2} C_{\text{opt}}\ .$$

Thus

$$\sum_{i=1}^{n} c_i x_i^d \geqslant \sum_{i=1}^{m} c_i y_i - \varepsilon C_{\text{opt}} = (1-\varepsilon) C_{\text{opt}}\ . \qquad \square$$

Lagrangian relaxation. Consider the (symmetric) travelling salesman problem again. For any vertex v, every tour uses two edges adjacent to v. Hence if we add the same constant to the length of every edge incident with v, we shift the value of every tour by the same number, and hence the problem remains essentially unchanged. By doing so for every vertex, we may bring the problem to a nicer form.

So far, this is the same idea as in the weighted bipartite matching algorithm above. Unfortunately, this does not lead to a complete solution; we cannot in general obtain an all-0 tour by shifting edge-weights like this. But we may use this method to improve dual heuristics. We have seen that a minimum length of a 1-tree is an easily computable lower bound; let us shift lengths so that this lower bound is maximized. This way we obtain a very good dual heuristic due to Held and Karp (1970).

It is not immediate how this new optimization problem can be solved. To describe a method, recall from matroid theory (chapter 11) that the convex hull of 1-trees with vertex set V is described by the constraints

$$\begin{aligned} &x_e \geqslant 0\,, \\ &x(A) \leqslant r(A)\,, \quad A \subseteq E\,, \\ &x(E) = n\,, \end{aligned} \tag{17}$$

Here $r(A)$ is the rank in the matroid whose bases are the 1-trees: if $c(A)$ is the number of connected components in (V, A), then

$$r(A) = \begin{cases} n - c(A) + 1\,, & \text{if } A \text{ contains a circuit}\,, \\ n - c(A) = |A|\,, & \text{if } A \text{ contains no circuits}\,. \end{cases}$$

If we want to restrict the feasible solutions to allow only tours, a natural step is to write up the degree constraints:

$$\sum_{j \in V \setminus \{i\}} x_{ij} = 2 \quad (i \in V)\,. \tag{18}$$

The objective function is

$$\text{minimize} \sum_e c(e) x_e\,.$$

The integral solutions of (17) and (18) are exactly the tours; if we drop the integrality constraints, we obtain a relaxation.

While it is trivial to minimize any objective function subject to constraints (17) using the greedy algorithm, constraints (18) spoil this nice structure. So let us get rid of the constraints (18) by multiplying them by appropriate multipliers λ_i, adding their left-hand sides to the objective function, and omitting them from the constraint set. Note that this leads to shifting the lengths of edges at nodes, as described above.

For any fixed choice of the multipliers, adding the left-hand sides of an equality constraint to the objective function does not change the problem (the objective function is shifted by the right-hand side); but then, dropping the constraint may decrease the optimum value. Can we choose the multipliers so that the optimum does not change? The answer is yes, and it is worth formulating the generalization of the Duality theorem of linear programming that guarantees this.

Theorem 9.5. *Consider a linear program with constraints split into two classes*:

$$\begin{aligned} \text{minimize} \quad & c^{\mathrm{T}} x \\ \text{subject to} \quad & Ax \geqslant a \\ & Bx \geqslant b\,. \end{aligned} \tag{19}$$

Then the optimum value of this program is the same as the optimum of the *following min–max problem*:

$$\max_y \{\min_x \{(c^{\mathrm{T}} - y^{\mathrm{T}}B)x \mid Ax \geq a\} + y^{\mathrm{T}}b,\ y \geq 0\}\ . \tag{20}$$

Similarly as in the Duality theorem, one can allow equations among the constraints, and then the corresponding multipliers y are unconstrained.

For any particular choice of the vector y, the minimum

$$\phi(y) := \min_x \{(c^{\mathrm{T}} - y^{\mathrm{T}}B)x \mid Ax \geq a\} + y^{\mathrm{T}}b$$

is a lower bound on the optimum value of (19). It is not difficult to see that the function ϕ, called the *Lagrange function* of (19), is a concave function, and hence various methods (subgradient, ellipsoid) are available to compute its maximum. Note that as long as we are only using this as a dual heuristic, we do not have to solve this problem to optimality: any reasonable y provides a lower bound.

Applying this technique to the travelling salesman problem, often rather good lower bounds are obtained. For example, the optimum value of the Lagrange function of the 52-city TSP of fig. 2.1 is equal to the length of the shortest tour.

It is worth mentioning that the Lagrangian relaxation method gives a result about exact solutions too.

Theorem 9.6. *Let $P \subseteq \mathbb{R}^n$ be a polytope and assume that every linear objective function $c^{\mathrm{T}}x$ ($c \in \mathbb{Z}^n$) can be minimized over P in time polynomial in $\langle c \rangle$. Then for every matrix $A \in \mathbb{Z}^{m \times n}$, and vectors $a \in \mathbb{Z}^m$ and $c \in \mathbb{Z}^n$, the minimum*

$$\min_x \{c^{\mathrm{T}}x \mid x \in P,\ Ax \leq a\}$$

can be computed in time polynomial in $\langle A \rangle + \langle a \rangle + \langle c \rangle$.

In other words, adding a few constraints to a nice problem does not spoil it completely. While this result could also be derived by other means (e.g., by the ellipsoid method), the Lagrangian approach is computationally much better if the number m of additional constraints is small.

10. Matrix methods

Determinants and matchings. Let G be a graph with n nodes. In chapter 3, a randomized algorithm (cf. Edmonds 1967b, and Lovász 1979) is described that decides if a graph G has a perfect matching. The method is based on the fact, proved by Tutte (1947) that det $A(G, x)$ is identically 0 if and only if G has no perfect matching, where $A(G, x)$ is the skew symmetric $n \times n$ matrix defined by

$$A(G,x)_{ij} = \begin{cases} x_{ij}\,, & \text{if } ij \in E(G) \text{ and } i < j\,,\\ -x_{ij}\,, & \text{if } ij \in E(G) \text{ and } i > j\,,\\ 0\,, & \text{otherwise}\,. \end{cases}$$

Generating random values for x_{ij}, and computing the determinant, we obtain a randomized matching algorithm. The following simple lemma, due to Schwartz (1980), can be used to estimate the probability of error.

Lemma 10.1. *Let $f(x_1, \ldots, x_p)$ be a polynomial, not identically* 0, *in which each variable has degree at most k. Choose the x_i independently from the uniform distribution on $\{0, 1, \ldots, N-1\}$. Then*

$$\mathrm{Prob}(f(x_1, \ldots, x_p) = 0) \leq \frac{k}{N}\,.$$

Since an $n \times n$ determinant can be evaluated in $\mathrm{O}(n^{2.39\cdots})$ time (Coppersmith and Winograd 1982), this randomized algorithm has a better running time than the best deterministic one, whose time complexity is $\mathrm{O}(n^{5/2})$ (Even and Kariv 1975).

We recall two variants of the determinant-based matching algorithm from chapter 3. The algorithm above determines whether a given graph has a perfect matching; but it does not give a perfect matching. To actually find a perfect matching, we can delete edges until we get a graph G_0 with a perfect matching such that deleting any further edge results in a graph with no perfect matching. Clearly, G_0 is a perfect matching itself.

Instead of this pedestrian procedure, Mulmuley et al. (1987) found an elegant randomized algorithm that finds a perfect matching at the cost of a single matrix inversion. The method is based on the following nice probabilistic lemma.

Lemma 10.2. *Let $(E, \mathcal{H})$ be a hypergraph and assign to each $e \in E$ a random weight w_e from the uniform distribution over $\{1, \ldots, 2|E|\}$. Then with probability at least $\frac{1}{2}$, the edge with minimum weight is unique.*

This lemma implies that if we substitute $x_{ij} = 2^{y_{ij}}$ in $A(G, x)$ (where each y_{ij} is uniformly chosen from $\{0, \ldots, 2n^2\}$) and then invert the resulting matrix A then, with probability at least $\frac{1}{2}$, those entries in the Schur product $A^{-1} \circ A$ having an odd numerator form a perfect matching. (The Schur product $C = A \circ B$ of two $n \times n$ matrices is defined by $C_{ij} = A_{ij}B_{ij}$.)

A special value of this method is that it is parallelizable, using polynomially many processors and polylog time. This depends on Csánky's theorem (see chapter 29, section 5) that gives an NC-algorithm for determinant computation and matrix inversion. Every known NC-algorithm for the perfect matching problem is randomized and uses determinant computation.

Given a graph $G = (V, E)$, an integer k, and $F \subseteq E$, the *exact matching problem* is to determine if there exists a perfect matching M in G such that $|M \cap F| = k$. This problem is not known to be in P, but it is easily solved in randomized polynomial time by the determinant method. Consider the matrix $A(G, x)$, and substitute yz_{ij} for x_{ij} if $ij \in F$, where y is a new variable. This way we obtain a matrix $A(G, z, y)$. Then Tutte's theorem on determinants and matchings can be extended as follows.

Theorem 10.1. *The coefficient of y^k in the Pfaffian* $\mathrm{Pf}(A(G, z, y))$ *is not identically* 0 *in the variables z iff there exists a perfect matching M such that $|M \cap F| = k$.*

This theorem suggests the following algorithm: Substitute random integers $\bar{z}_{ij} \in \{0, \ldots, N-1\}$ in $A(G, y, z)$. The value of det $A(G, y, \bar{z})$ is a polynomial in y, and all its coefficients can be computed in polynomial time. Compute $\mathrm{Pf}(A(G, y, \bar{z})) = \sqrt{\det A(G, y, \bar{z})}$, which is also a polynomial in y by definition. The coefficient of y^k gives the answer. (See chapter 3, section 7 for other applications of this idea.)

Determinants and connectivity. The method of reducing a combinatorial optimization problem to checking a polynomial (usually determinantal) identity and then solving this in randomized polynomial time via Schwartz's lemma 10.1 is not restricted to matching theory. Chapter 36 contains examples where this method is used in electrical engineering and statics. The papers Linial et al. (1988) and Lovász et al. (1989) contain various algorithms to determine the connectivity of a graph along these lines. Let us formulate one of these. Let G be a graph and consider, for each (unordered) pair ij with $i = j$ or $ij \in E(G)$, a variable x_{ij}. Let $B(G, x)$ be the matrix

$$B(G,x)_{ij} = \begin{cases} x_{ij}, & \text{if } i = j \text{ or } ij \in E(G), \\ 0, & \text{otherwise}. \end{cases}$$

Theorem 10.2. *The graph G is k-connected iff*

$$\text{no } (n-k) \times (n-k) \text{ subdeterminant of } B(G, x) \text{ is identically } 0. \qquad (*)$$

This theorem suggests the following randomized k-connectivity test: substitute in $B(G, x)$ independent random integers from, say, $\{0, \ldots, 2^n\}$, and check condition (*). Unfortunately, there is no polynomial time algorithm known to check (*) for a general matrix; however, due to the very special structure of $B(G, x)$, it suffices to check only a "few" subdeterminants. Let us select, for each vertex i, a set A_i of $k-1$ neighbors (if a node has degree less than $k-1$ then the graph is clearly not k-connected). Then the following can be shown: if G is not k-connected, then one of the subdeterminants of $B(x)$, obtained by deleting the rows belonging to some $A_i \cup \{i\}$ and the columns belong to some $A_j \cup \{j\}$, is identically 0.

This leads to the evaluation of $O(n^2)$ determinants of size $(n-k) \times (n-k)$. With a little care, one can reduce this number to $O(nk)$. For $k < n/2$, it is worth inverting the matrix $B(G, x)$ and then check $O(nk)$ subdeterminants of size $k \times k$ using Jacobi's theorem (see chapter 31).

Semidefinite optimization. Polyhedral combinatorics can be viewed as a theory of linear inequalities valid for the incidence vectors of various set-systems. It is quite natural to ask for quadratic inequalities (and, of course higher degree inequalities) valid for these incidence vectors. This idea leads to real algebraic geometry and its study has just begun.

At first sight it seems that we are getting too much too easily. Let $G = (V, E)$ be a graph, $V = \{1, \dots, n\}$, and consider the following system of equations:

$$x_i^2 = x_i \quad \text{for every node } i \in V\,, \tag{21}$$

$$x_i x_j = 0 \quad \text{for every edge } ij \in E\,. \tag{22}$$

Trivially, the solutions of (21) are precisely the 0–1 vectors, and so the solutions of (21)–(22) are precisely the incidence vectors of stable sets. Unfortunately, there is little known about the solutions of systems of quadratic equations. In fact, this construction shows that even the solvability of such a simple system of quadratic equations (together with a linear equation $\sum_i x_i = \alpha$) is NP-hard.

However, we can use this system to derive some other constraints. (21) implies that for every node i,

$$x_i = x_i^2 \geqslant 0\,, \qquad 1 - x_i = (1 - x_i)^2 \geqslant 0\,. \tag{23}$$

Using this, (22) implies that for every edge ij,

$$1 - x_i - x_j = 1 - x_i - x_j + x_i x_j = (1 - x_i)(1 - x_j) \geqslant 0\,. \tag{24}$$

So we can derive the edge constraints from (21)–(22) formally. We can also derive the clique constraints. Assume that nodes $1, \dots, k$ induce a complete subgraph. We start with the trivial inequality

$$(1 - x_1 - \cdots - x_k)^2 \leqslant 0\,.$$

Expanding, we get

$$1 + \sum_{i=1}^{k} x_i^2 - 2 \sum_{i=1}^{k} x_i + 2 \sum_{i \neq j} x_i x_j \geqslant 0\,.$$

Here the first sum is just $\sum_i x_i$ by (21) and the third sum is 0 by (22), so we get

$$1 - x_1 - \cdots - x_k \geqslant 0\,. \tag{25}$$

In the special case when the graph is perfect, we obtain all constraints for the stable set polytope STAB(G) in a single step. (STAB(G) is the convex hull of all incidence vectors of stable sets of G.)

The algorithmic significance of this observation is that it leads to a polynomial time algorithm to compute the stability number of a perfect graph. By general consequences of the ellipsoid method (see chapter 30), it suffices to design a polynomial time algorithm that checks whether a given inequality

$$\sum_i c_i x_i \leqslant \gamma \tag{26}$$

is valid for STAB(G). By the above arguments, it suffices to check whether (26) can be derived from (21) and (22) as above. Formalizing, this leads to the question: do there exist real multipliers μ_i $(i \in V)$ and λ_{ij} $(ij \in E)$, and linear polynomials $l_1, \ldots, l_m$ such that

$$\sum_{k=1}^{m} l_k^2 + \sum_{i=1}^{n} \mu_i(x_i^2 - x_i) + \sum_{ij \in E} \lambda_{ij} x_i x_j = \gamma - \sum_{i=1}^{n} c_i x_i .$$

This is equivalent to saying that there exist λ's and μ's such that the $(n+1) \times (n+1)$ matrix $P = (p_{ij})$ defined by

$$p_{ij} := \begin{cases} \gamma , & \text{if } i = j = 0 , \\ (\mu_i - c_i)/2 , & \text{if } j = 0, i > 0 , \\ (\mu_j - c_j)/2 , & \text{if } i = 0, j > 0 , \\ -\mu_i , & \text{if } i = j > 0 , \\ \lambda_{ij}/2 , & \text{if } ij \in E , \\ 0 , & \text{otherwise} , \end{cases}$$

is positive semidefinite.

More generally, we can consider optimization problems of the type

$$\begin{aligned} &\text{maximize} \quad \sum_i c_i y_i \\ &\text{subject to} \quad P(y) \text{ is positive definite} , \end{aligned} \tag{27}$$

where $P(y)$ is a matrix in which every entry is a linear function of the y_i. Grötschel et al. (1981) describe a way, using the ellipsoid method, to solve such problems to arbitrary precision in polynomial time. (The key facts are that the feasible domain is convex and to check whether a given x satisfies the constraints can be done using Gaussian elimination; we ignore numerical difficulties here that are non-trivial.) The Duality theorem can also be extended to such semidefinite programs (see Wolkowitz 1981). Recently Alizadeh (1992) showed that Karmarkar's interior point method can also be extended to such semidefinite programs in a very natural way, which is much more promising from a practical point of view.

The method sketched here can be used to generate other classes of inequalities for STAB(G) and to show their polynomial time solvability. It is not restricted to the stable set problem either; in fact, it can be applied to any 0–1 optimization problem. Interesting applications of a related method to the max-cut problem were given by Delorme and Poljak (1990) (see also Mohar and Poljak 1990).

Both of these semidefinite relaxations can be combined with randomized rounding methods (section 8). This results in interesting approximation algorithms for the max-cut problem (Goemans and Williamson 1994) and for chromatic number (Karger et al. 1994).

These methods can be extended from quadratic to higher-order inequalities.

For these extensions, see Lovász and Schrijver (1990) and Sherali and Adams (1990).

References

Alizadeh, F.
[1992] Combinatorial optimization with semi-definite matrices, in: *Integer Programming and Combinatorial Optimization, Proc. IPCO '92*, eds. E. Balas, G. Cornuéjols and R. Kannan (Carnegie Mellon University Printing, Pittsburgh, PA) pp. 385–405.

Alon, N., R.A. Duke, H. Lefmann, V. Rödl and R. Yuster
[1994] The algorithmic aspects of the Regularity Lemma, *J. Algorithms* **16**, 80–109.

Applegate, D., and R. Kannan
[1990] Sampling and integration of near logconcave functions, in: *Proc. 23th ACM STOC*, pp. 156–163.

Applegate, D., R. Bixby, V. Chvátal and W. Cook
[1994] *Finding Cuts in the TSP*, Preliminary report.

Bachem, A., M. Grötschel and B. Korte
[1983] *Mathematical Programming: The State of the Art* (Springer, Heidelberg).

Barahona, F., and A.R. Mahjoub
[1986] On the cut polytope, *Math. Programming* **36**, 157–173.

Barahona, F., M. Grötschel, M. Jünger and G. Reinelt
[1988] An application of combinatorial optimization to statistical physics and circuit layout design, *Oper. Res.* **36**, 493–513.

Bellman, R.
[1957] *Dynamic Programming* (Princeton University Press, Princeton, NJ).

Bentley, J.L.
[1992] Fast algorithms for geometric traveling salesman problems, *ORSA J. Comput.* **4**, 387–411.

Christofides, N.
[1976] *Worst-case Analysis of a new Heuristic for the Travelling Salesman Problem*, Technical Report (Graduate School of Industrial Adminstration, Carnegie-Mellon University, Pittsburgh, PA).

Chvátal, V., and G. Klincsek
[1980] Finding largest convex subsets, *Congress. Numerantium* **29**, 453–460.

Coffman Jr, E.G., M.R. Garey and D.S. Johnson
[1984] Approximation algorithms for bin-packing – an updated survey, in: *Algorithm Design for Computer System Design*, eds. G. Ausiello, M. Lucertini and P. Serafini (Springer, New York) pp. 49–106.

Coppersmith, D., and S. Winograd
[1982] On the asymptotic complexity of matrix multiplication, *SIAM J. Comput.* **11**, 472–492.

Cormen, T.H., C.E. Leiserson and R.L. Rivest
[1990] *Introduction to Algorithms* (MIT Press, Cambridge, MA).

Delorme, C., and S. Poljak
[1993] Laplacian eigenvalues and the maximum cut problem, *Math. Programming A* **62**, 557–574.

Deza, M., and M. Laurent
[1991] *A Survey of the known Facets of the Cut Cone*, Report No. 91722-OR (Forschungsinstitut für Diskrete Mathematik, Universität Bonn).

Dyer, M., and A. Frieze
[1992] Computing the volume of convex bodies: A case where randomness provably helps, in: *Probabilistic Combinatorics and Its Applications*, ed. B. Bollobás, *Proc. Symp. Appl. Math.* **44**, 123–170.

Dyer, M., A. Frieze and R. Kannan
[1991] A random polynomial time algorithm for approximating the volume of convex bodies, *J. Assoc. Comput. Mach.* **38**, 1–17.

Edmonds, J.
[1967a] Optimum branchings, *J. Res. Nat. Bur. Standards B* **71**, 233–240.

Edmonds, J., and R.M. Karp

[1972] Theoretical improvements in algorithmic efficiency for network flow problems, *J. Assoc. Comput. Mach.* **19**, 248–264.

Edmonds, J.R.

[1967b] Systems of distinct representatives and linear algebra, *J. Res. Nat. Bur. Standards B* **71**, 241–247.

Erickson, R.E., C.L. Monma and A.F. Veinott

[1987] Send-and-split method for minimum concave-cost network flows, *Math. Oper. Res.* **12**, 634–664.

Even, S., and O. Kariv

[1975] An O($n^{5/2}$) algorithm for maximum matching in general graphs, in: *16th Annu. Symp. on Foundations of Computer Science* (IEEE Computer Society Press, New York) pp. 100 112.

Fernandez de la Vega, W., and G.S. Lueker

[1981] Bin packing solved within $1+\varepsilon$ in linear time, *Combinatorica* **1**, 349–355.

Ford Jr, L.R., and D.R. Fulkerson

[1962] *Flows in Networks* (Princeton University Press, Princeton, NJ).

Frank, A., and É. Tardos

[1987] An application of simultaneous Diophantine approximation in combinatorial optimization, *Combinatorica* **7**, 49–65.

Garey, M.R., R.L. Graham, D.S. Johnson and A. Yao

[1976] Resource constrained scheduling as generalized bin packing, *J. Combin. Theory A* **21**, 257–298.

Goemans, M.X., and D.P. Williamson

[1994] .878-Approximation algorithms for MAX CUT and MAX 2SAT, in: *Proc. 26th ACM Symp. on Theory of Computing* (ACM, New York) pp. 422–431.

Goldberg, M.V., É. Tardos and R.E. Tarjan

[1990] Network flow algorithms, in: *Paths, Flows, and VLSI-Layout*, eds. B. Korte, L. Lovász, H.J. Prömel and A. Schrijver (Springer, Heidelberg) pp. 101–164.

Gondran, M., and M. Minoux

[1979] *Graphes et Algorithmes* (Editions Eyrolles, Paris).

Graham, R.L., and P. Hell

[1985] On the history of the minimum spanning tree problem, *Ann. Hist. Comput.* 7, 43–57.

Grötschel, M., and O. Holland

[1991] Solution of large-scale symmetric travelling salesman problems, *Math. Programming* **51**, 141–202.

Grötschel, M., and M.W. Padberg

[1985] Polyhedral theory, in: *The Travelling Salesman Problem: A Guided Tour through Combinatorial Optimization*, eds. E.L. Lawler, J.K. Lenstra, A.H.G. Rinnooy Kan and D.B. Shmoys (Wiley, Chichester) pp. 251–305.

Grötschel, M., and Y. Wakabayashi

[1989] A cutting plane algorithm for a clustering problem, *Math. Programming B* **45**, 59–96.

Grötschel, M., L. Lovász and A. Schrijver

[1981] The ellipsoid method and its consequences in combinatorial optimization, *Combinatorica* **1**, 169–197.

Grötschel, M., M. Jünger and G. Reinelt

[1984] A cutting plane algorithm for the linear ordering problem, *Oper. Res.* **32**, 1195–1220.

Grötschel, M., L. Lovász and A. Schrijver

[1988] *Geometric Algorithms and Combinatorial Optimization* (Springer, Heidelberg).

Grötschel, M., A. Martin and R. Weismantel

[1992a] Packing Steiner trees: a cutting plane algorithm and computational results, Preprint 92–9 (Konrad-Zuse-Zentrum für Informationstechnik Berlin), *Math. Programming*, to appear.

Grötschel, M., C. Monma and M. Stoer

[1992b] Computational results with a cutting plane algorithm for designing communication networks with low connectivity constraints, *Oper. Res.* **40**, 309–330.

Haken, A., and M. Luby

[1988] Steepest descent can take exponential time for symmetric connection networks, *Complex Systems* **2**, 191–196.

Haussler, D., and E. Welzl

[1987] ε-nets and simplex range queries, *Discrete Comput. Geom.* **2**, 127–151.

Held, M., and R.M. Karp

[1970] The travelling salesman problem and minimum spanning trees, *Oper. Res.* **18**, 1138–1162.

Hoffman, K.L., and M. Padberg

[1993] Solving airline crew-scheduling problems by branch-and-cut, *Manage. Sci.* **39**, 657-682.

Holley, R., and D. Stroock

[1988] Simulated annealing via Sobolev inequalities, *Comm. Math. Phys.* **115**, 553–569.

Holley, R., S. Kusuoka and D. Stroock

[1989] Asymptotics of the spectral gap with applications to the theory of simulated annealing, *J. Funct. Anal.* **83**, 333–347.

Hurkens, C.A.J., L. Lovász, A. Schrijver and É. Tardos

[1988] How to tidy up your set-system? in: *Combinatorics, Proc. Coll. Eger 1987,* eds. L. Lovász and V.T. Sós, *Colloq. Math. Soc. János Bolyai* **52**, 309 314.

Ibarra, O.H., and C.E. Kim

[1975] Fast approximation algorithms for the knapsack and sum of subset problems, *J. Assoc. Comput. Mach.* **22**, 463 468.

Jenkyns, T.

[1976] The efficacy of the "greedy" algorithm, in: *Proc. 7th South-Eastern Conf. on Combinatorics, Graph Theory and Computing* (Utilitas Math., Winnipeg) pp. 341–350.

Jerrum, M.R.

[1992] Large cliques elude the Metropolis process, *Rand. Struetures Algorithms* **3**, 347–359.

Jerrum, M.R., and A.J. Sinclair

[1989] Approximating the permanent, *SIAM J. Comput.* **18**, 1149–1178.

Jerrum, M.R., L.G. Valiant and V.V. Vazirani

[1986] Random generation of combinatorial structures from a uniform distribution, *Theor. Comput. Sci.* **43**, 169–188.

Johnson, D.S.

[1973] *Near-optimal allocation algorithms,* Ph.D. Dissertation (MIT, Cambridge, MA).

[1974] Approximation algorithms for combinatorial problems, *J. Comput. System. Sci.* **9**, 256–298.

[1990] Local optimization and the travelling salesman problem, in: *Proc. 17th Coll. on Automata, Languages and Programming* (Springer, Heidelberg) 446 461.

Johnson, D.S., C.H. Papadimitriou and M. Yannakakis

[1988] How easy is local search? *J. Comput. System Sci.* **37**, 79–100.

Johnson, D.S., C.R. Aragon, L.A. McGeoch and C. Schevon

[1989] Optimization by simulated annealing: an experimental evaluation; Part I, graph partitioning, *Oper. Res.* **37**, 865–892.

[1991] Optimization by simulated annealing: an experimental evaluation; Part II, graph coloring and number partitioning, *Oper. Res.* **39**, 378–406.

Jünger, M., G. Reinelt and G. Rinaldi

[1995] The travelling salesman problem, in: *Network Models, Handbooks in Operations Research and Management Science,* Vol. 7, eds. M.O. Ball, T.L. Magnanti, C.L. Monma and G.L. Nemhauser (North-Holland, Amsterdam).

Karger, D., R. Motwani and M. Sudan

[1994] Approximate graph coloring by semidefinite programming, in: *35th Annu. Symp. on Foundations of Computer Science* (IEEE Computer Society Press, New York) to appear.

Karmarkar, N., and R.M. Karp

[1982] An efficient approximation scheme for the one-dimensional bin-packing problem, in: *23rd Annu. Symp. on Foundations of.Computer Science* (IEEE Computer Society Press, New York) pp. 312 320.

Kirckpatrick, S., C.D. Gelatt and M.P. Vecchi

[1983] Optimization by simulated annealing, *Science* **230**, 671–680.

Komlós, J., J. Pach and G. Woeginger

[1992] Almost tight bounds on epsilon-nets, *Discrete Comput. Geom.* 7, 163 173.

Korte, B., and D. Hausmann

[1978] An analysis for the greedy algorithm for independence systems, *Ann. Discrete Math.* **2**, 65–74.

Korte, B., L. Lovász and R. Schrader

[1991] *Greedoids* (Springer, Heidelberg).

Lawler, E.L., J.K. Lenstra, A.H.G. Rinnooy Kan and D.B. Shmoys

[1985] eds., *The Travelling Salesman Problem: A Guided Tour through Combinatorial Optimization* (Wiley, Chichester).

Lenstra, J.K., D.B. Shmoys and É. Tardos

[1990] Approximation algorithms for scheduling unrelated parallel machines, *Math. Programming A* **46**, 259–271.

Lin, S., and B.W. Kernighan

[1973] An effective heuristic algorithm for the traveling salesman problem, *Oper. Res.* **21**, 498–516.

Linial, N., L. Lovász and A. Wigderson

[1988] Rubber bands, convex embeddings, and graph connectivity, *Combinatorica* **8**, 91–102.

Lovász, L.

[1975] On the ratio of optimal fractional and integral covers, *Discrete Math.* **13**, 383–390.

[1979] Determinants, matchings, and random algorithms, in: *Fundamentals of Computation Theory, FCT'79*, ed. L. Budach (AkademieVerlag, Berlin) pp. 565–574.

Lovász, L., and M.D. Plummer

[1986] *Matching Theory, Ann. Discrete Math.* **29**.

Lovász, L., and A. Schrijver

[1990] Cones of matrices and setfunctions, and 0–1 optimization, *SIAM J. Optim.* **1**, 166–190.

Lovász, L., and M. Simonovits

[1992] On the randomized complexity of volume and diameter, *Proc. 33rd IEEE Annu. Symp. on Foundations of Computer Science* (IEEE Computer Society Press, New York) pp. 482–491.

Lovász, L., M. Saks and A. Schrijver

[1989] Orthogonal representations and connectivity of graphs, *Linear Algebra Appl.* **114/115**, 439–454.

Metropolis, N., A. Rosenblut, M. Rosenbluth, A. Teller and E. Teller

[1953] Equation of state calculation by fast computing machines, *J. Chem. Phys.* **21**, 1087–1092.

Mohar, B., and S. Poljak

[1990] Eigenvalues and the max-cut problem, *Czech. Math. J.* **40**, 343–352.

Monge, G.

[1781] *Déblai et remblai*, Memoires de l'Academie des Sciences.

Mulmuley, K., U.V. Vazirani and V.V. Vazirani

[1987] Matching is as easy as matrix inversion, *Combinatorica* **7**, 105–113.

Nemhauser, G.L., and L.E. Trotter Jr

[1974] Properties of vertex packing and independence system polyhedra, *Math. Programming* **6**, 48–61.

Nemhauser, G.L., and L.A. Wolsey

[1988] *Integer and Combinatorial Optimization* (Wiley, Chichester).

Nemhauser, G.L., A.H.G. Rinnooy Kan and M.J. Todd

[1989] *Optimization, Handbooks in Operations Research and Management Science*, Vol. 1 (North-Holland, Amsterdam).

Padberg, M.W., and M. Grötschel

[1985] Polyhedral computations, in: *The Travelling Salesman Problem: A Guided Tour through Combinatorial Optimization*, eds. E.L. Lawler, J.K. Lenstra, A.H.G. Rinnooy Kan and D.B. Shmoys (Wiley, Chichester) pp. 307–360.

Padberg, M.W., and G. Rinaldi

[1991] A branch-and-cut algorithm for the solution of large-scale traveling salesman problems, *SIAM Rev.* **33**, 1–41.

Poljak, S.

[1994] Integer linear programs and local search for max-cut, *SIAM J. Comput.*, to appear.

Raghavan, P., and C.D. Thompson

[1987] Randomized rounding: a technique for provably good algorithms and algorithmic proofs, *Combinatorica* **7**, 365–374.

Reinelt, G.

[1994] *The Traveling Salesman: Computational Solutions for TSP Applications, Lecture Notes in Computer Science*, Vol. 840 (Springer, Heidelberg).

Rosenkrantz, D.J., R.E. Stearns and P.M. Lewis II

[1977] An analysis of several heuristics for the traveling salesman problem, *SIAM J. Comput.* **6**, 563–581.

Sasaki, G.

[1991] The effect of the density of states on the Metropolis algorithm, *Inform. Process. Lett.* **37**, 159–163.

Sasaki, G.H., and B. Hajek

[1988] The time complexity of maximum matching by simulated annealing, *J. Assoc. Comput. Mach.* **35**, 387–403.

Schäffer, A., and M. Yannakakis

[1991] Simple local search problems that are hard to solve, *SIAM J. Comput.* **20**, 56–87.

Schrijver, A.

[1986] *Theory of Integer and Linear Programming* (Wiley, Chichester).

Schwartz, J.T.

[1980] Fast probabilistic algorithms for verification of polynomial identities, *J. Assoc. Comput. Mach.* **27**, 701–717.

Sherali, H.D., and W.P. Adams

[1990] A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems, *SIAM J. Discrete Math.* **3**, 411–430.

Shmoys, D.B., and É. Tardos

[1993] An approximation algorithm for the generalized assignment problem, *Math. Programming A* **62**, 461–474.

Sinclair, A.J., and M.R. Jerrum

[1988] Conductance and the rapid mixing property for Markov chains: The approximation of the permanent resolved, in: *Proc. 20th ACM STOC* (ACM, New York) pp. 235–244.

Stein, S.K.

[1974] Two combinatorial covering theorems, *J. Combin. Theory A* **16**, 391–397.

Tutte, W.T.

[1947] The factorisation of linear graphs, *J. London Math. Soc.* **22**, 107–111.

[1963] How to draw a graph, *Proc. London Math. Soc.* **13**, 743–768.

van Vliet, A.

[1992] An improved lower bound for on-line bin-packing algorithms, *Inform. Process. Lett.* **43**, 277–284.

Vapnik, V.N.

[1982] *Estimation of Dependences Based on Empirical Data* (Springer, New York).

Wolkowitz, H.

[1981] Some applications of optimization in matrix theory, *Linear Algebra Appl.* **40**, 101–118.

CHAPTER 29

Computational Complexity

D.B. SHMOYS and É. TARDOS

School of Operations Research and Industrial Engineering, Upson Hall, Cornell University, Ithaca, NY 14853, USA

Contents

HANDBOOK OF COMBINATORICS
Edited by R. Graham, M. Grötschel and L. Lovász

Computational complexity theory attempts to understand the power of computation by providing insight into the question as to why certain computational problems appear to be more difficult than others. Computation has added a dimension to the study of combinatorics. The theorem that, given a matching in a graph, *there exists* a larger matching if and only if there is an augmenting path, is not the complete answer; is it possible to *efficiently construct* a larger matching if one exists? Although such an algorithm is known for the matching problem, this is not the case for many combinatorial problems. Indeed, the greatest challenge confronting complexity theory is to provide techniques to prove that no efficient algorithm exists for a given problem.

Computational complexity theory provides the mathematical framework in which to discuss these questions, and while substantial progress has been made towards distinguishing the difficulty of computational problems, most of the basic issues remain unresolved. In this chapter, we will describe the fundamentals of this theory and give a brief survey of the results that have been obtained in its first quarteŕ-century. For a more detailed and complete exposition, the reader is referred to the textbooks by Garey and Johnson (1979) and Hopcroft and Ullman (1979) as well as to the more recent *Handbook of Theoretical Computer Science* edited by van Leeuwen (1990).

1. Complexity of computational problems

In this section, we will outline the essential machinery used to give formal meaning to the complexity of computational problems. This involves describing what precisely is meant by a computational problem, setting up a mathematical model of computation, and then formalizing the notion of the computational resources required for a problem with respect to that model. Unfortunately, there is no one standardized specification under which to discuss these questions. For this theory to produce meaningful results, it is essential that the definitions be robust enough that theorems proved with respect to them apply equally to all reasonable variants of this framework. Indeed, the particular definitions that we will rely on will be accompanied by evidence that these notions are sufficiently flexible.

1.1. Computational problems

Computation can be thought of as finding a suitable output for a given input. Therefore, a *computational problem* is specified by a relation between inputs and output; an algorithm to solve the problem takes an acceptable input, called an *instance*, and computes an output that satisfies the input – output relation; for example, given a directed graph, output a hamiltonian circuit, if there is one, and otherwise indicate that none exists. This framework is very general, and we will focus attention on certain sorts of inputs and outputs.

The most common type of input (or output) is a string of characters over a finite alphabet. The previous example can be cast in this setting, since it is easy to represent a graph of order n as such a string of 0's and 1's; one might concatenate the rows of the $n \times n$ matrix $A = (a_{ij})$ where $a_{ij} = 1$ if and only if (i, j) is an arc. Alternatively, one might list the names of the nodes incident from node i, for $i = 1, \ldots, n$, where the node named j is encoded by the binary representation of j. Observe that the second representation will be more compact if the graph has few arcs, but this difference is limited, in that the length of the input in one format is at most the square of the length in the other.

Some combinatorial structures cannot be compactly represented as a string; for example, a matroid on n elements is most naturally represented by a string of length 2^n, where each character indicates whether a particular subset is independent. In such cases, it is customary to use an *oracle* to specify the input; the algorithm may write down queries, such as a particular subset to be tested for independence, and the oracle's answer may be used in the next step of the algorithm. Sometimes, there is no natural finite representation of the input, such as in the problem of optimizing over a convex body. For this example, one standard way to give the input is via an oracle that decides whether a given point is in the body, and if not, outputs a separating hyperplane. For any oracle, there is an associated parameter which is used as a measure of the size of the input.

In analyzing algorithms, we often view numbers as atomic units, without regard for their lengths, and so an input can also be a list of numbers, where the size of the input is the number of elements in the list. However, for the remainder of this chapter we will focus on inputs that are strings, although it is straightforward to extend the discussion of computational models and complexity to include these alternatives.

An important special case of a computational problem is a *decision problem*, where the output is restricted to either "yes" or "no", and for each input, there is exactly one related output. The set L of "yes" instances for a decision problem is often called the *language* associated with this problem. A decision version of the previous example is: given a directed graph, does it contain a hamiltonian circuit? This type of problem will play a central role in our discussion, and it is important to realize that it is not a significant limitation to focus on it. For example, it is possible to answer the first *search* version of the hamiltonian circuit problem as follows: for each arc in the graph, delete the arc, decide if the resulting graph is still hamiltonian and replace the arc only in the case that the answer is "no". At the end of this procedure, the remaining graph is a circuit. Thus, we have shown that finding a circuit is not much harder than the decision problem, and this relationship remains true for all well-formulated decision versions of search problems.

Another important type of computational problem is an *optimization problem*; for example, given a directed graph G and two nodes s and t, we may wish to find the shortest path from s to t (or merely find the length). We shall in fact treat these as decision problems by adding a bound b to the instance, and, for example, asking whether G has a path from s to t of length at most b. By using a *binary search* procedure that iteratively halves the range of possible optimal values, we see that an optimization problem can be solved with only somewhat more work than the corresponding decision problem.

Throughout this chapter, we will be dealing with computational problems involving a variety of structures, and we will not be specifying the nature of the encodings used. We will operate on the premise that any reasonable encoding produces strings of length that can be bounded by a polynomial of the length produced by any other encoding. When encoding numbers (which we will assume to be integral) there is an important distinction between the binary representation, which we will typically use, and the unary representation (e.g., representing 5 as 11111). Notice that the latter could be exponentially bigger than the former, and thus size is a deceptive measure, since it makes instances of the problem larger than they need be. As we survey the complexity of computational problems that involve numbers, we will see that some are sensitive to this choice of encodings, whereas others are less affected.

1.2. Models of computation and computability

We next turn our attention to defining a mathematical model of a computer. In fact, we will present three different models, and although their superficial characteristics make them appear quite different, they will turn out to be formally equivalent. The first of these is the *Turing machine*, which is an extremely primitive model, and as a result, it is easier to prove results about what cannot be computed within this model. On the other hand, its extreme simplicity makes it ill-suited for algorithm design. As a result, it will be convenient to have an alternative model, the random access machine (RAM), within which to discuss algorithms.

The name "Turing machine" is a slight misnomer, since a Turing machine is a mathematical formulation of an algorithm, rather than a machine. A Turing machine $M = (Q, \Gamma, \delta, q_0, A)$ is a machine that has a finite main memory represented by a finite set of states Q, a read-only input tape, a finite set of work tapes each of which contains a countably infinite number of cells (corresponding to the integers) to store a character from a finite alphabet Γ, which at least includes the input alphabet $\{0, 1\}$ and a blank symbol B. For each of the k work tapes and the input tape, there is a "head" that can read one cell of the tape at a given time, and will be able to move cell-by-cell across the tape as the computation proceeds. Throughout the computation, the heads will read the contents of cells, and depending on what was read and the current state of the main memory, rewrite the cells and then move each head by one cell, either left or right, as well as cause a change in the state of the main memory. The basic step of a Turing machine, a *transition*, is modeled by a partial function

$$\delta : Q \times \Gamma^{k+1} \mapsto Q \times \Gamma^{k} \times \{L, R\}^{k+1}$$

that selects the new state, the contents of the cells currently scanned on the work tapes, and indicates the direction in which each head moves one cell as a *deterministic* function of the current state and the contents of the input- and work-tape cells currently being read. One may view the transition function as the program hardwired into this primitive machine. The computation is begun in state q_0, with the input head at the left end of the input, and all of the work tapes and the rest of the input tape blank. The machine *halts* if δ is undefined for the current state and the symbols read. An input is *accepted* if it halts in a state in the accepting state set $A \subseteq Q$. A Turing machine M solves a decision problem L if L is the set of inputs accepted by M, and M halts on every input; such a language L is said to be *decidable*. This definition of a Turing machine is similar to the one given by Turing (1936), and the reader should note that many equivalent definitions are possible.

We have defined a Turing machine so that it can only solve decision problems, but this definition can be easily extended to arbitrary computational problems by, for example, adding a write-only output tape, on which to print the output before halting. Although this appears to be a very primitive form of a computer, it has become routine to accept the following proposition.

Church's thesis: *Any function computed by an effective procedure can be computed by a Turing machine.*

Although this is a *thesis*, in the sense that any attempt to characterize the inexplicit notion of effective procedure would destroy its intent, it is supported by a host of *theorems*, since for any known characterization of computable functions, it has been shown that these are Turing computable.

A *random access machine* (RAM) is a model of computation that is well-suited to specifying algorithms, since it uses an idealized, simplified programming language that closely resembles the assembly language of any modern-day digital computer. There is an infinite number of memory cells indexed by the integers, and there is no bound on the size of an integer that can be stored in any cell. A program can directly specify a cell to be read from or written in, without moving a head into position. Furthermore, there is an indirect addressing option, which uses the contents of a cell as an index to another cell that is then (seemingly randomly) accessed. All basic arithmetic operations can be performed. For further details on RAM's the reader is referred to Aho, Hopcroft and Ullman (1974).

Another model of computation closely tied to a practical setting is the *logical circuit model*. The simplicity of the circuit model makes it extremely attractive for proving lower bounds on the computational resources needed for particular functions, and research along these lines will be discussed in depth in chapter 40. A circuit may be thought of as a directed acyclic graph, where the nodes of indegree 0 are the Boolean *input gates* (which can assume value 0 or 1), and the remainder correspond to *functional gates* of the circuit and are labeled with an operation, such as the logical or, negation, or logical and operations. The nodes of outdegree 0 are the *outputs*. A given circuit only handles inputs of a particular size, and so we specify a circuit for each input length. We say that the family of circuits solves a computational problem if for each input the corresponding circuit in the family generates a related output. Note that this model differs from the previous two in that the circuit for inputs of length n can be tailored *nonuniformly* to the particular value of n, whereas a Turing machine or RAM must run on inputs *uniformly*, independent of their length. We can make the notion of a family of circuits equivalent to the other models of computation by insisting that there be a Turing machine that, on input n, computes the description of the circuit for inputs of length n.

In spite of the apparent differences in these three models, any particular choice is one of convenience, and not of substance.

Theorem 1.1. *The following classes of problems are identical:*

- *the class of computational problems solvable by a Turing machine;*
- *the class of computational problems solvable by a RAM;*
- *the class of computational problems solvable by a family of circuits that can be generated by a Turing machine.*

Theorem 1.1 is complemented by the following theorem, which is a consequence of the fact that there are an uncountable number of decision problems, but only a countable number of Turing machines.

Theorem 1.2. *Not all decision problems are solvable by a Turing machine.*

The understanding of this inherent limitation on the power of computation was an outgrowth of results in mathematical logic. In particular, the first incompleteness theorem of Gödel (1931) contained the first sort of undecidability result and provided many of the essential ideas that would be used by Church, Post and Turing in their groundbreaking work on the nature of computation. In particular, Turing (1936) proposed what we call a Turing machine, and this enabled the discussion to be conveniently directed towards computation.

At the core of all of these results is Gödel's notion of encoding theorems as strings in some uniform way. Analogously, a Turing machine can be encoded as a string by first specifying the number of states that the machine has, followed by a list of all of the allowed transitions. Each such string can also be interpreted as an integer by using a binary encoding. Thus, each integer i represents a Turing machine M_i. Turing showed that the language $L_{\mathrm{hp}} = \{i \mid M_i \text{ halts on input } i\}$ is not solvable by a Turing machine. His proof that this *halting problem* is undecidable uses the following diagonalization argument. Suppose that L_{hp} were solvable by a Turing machine M. Build another machine M' that first uses M to decide if the input $i \in L_{\mathrm{hp}}$; if it is, then M' enters an infinite loop, and if not, M' halts. Of course, M' must be M_k for some integer k. But M' and M_k cannot accept the same language, since M' halts on k if M_k does not, and vice versa.

A problem L_1 (many-one) *reduces* to a problem L_2 if there exists a function f computable by a Turing machine such that $x \in L_1$ if and only if $f(x) \in L_2$. Note that if L_1 is undecidable and L_1 reduces to L_2, then L_2 must also be undecidable. This provides a strategy for proving additional undecidable problems. As a simple example, consider the language $L_{\mathrm{e}} = \{i \mid M_i \text{ accepts on an empty input}\}$. It is a simple exercise to convert the description of a given Turing machine M_i to the description of another (rather trivial) machine $M' = M_j$ that on every input, first runs M_i with input i and accepts if M_i halts on i. Clearly, M_j accepts an empty input if and only if M_i halts on i.

A similar strategy can be used to prove Gödel's undecidability theorem in the context of Turing computability, although the details of the reduction are more involved. The theory of arithmetic for non-negative integers with addition and multiplication can be defined as follows. Consider first-order formulas that can be constructed from variables and the constants 0 and 1 with the logical connectives $\neg$, $\vee$, $\wedge$, $\to$, $\exists$, and $\forall$, along with the operations $\cdot$ and $+$, and the binary relations $=$ and $<$. A sentence is a formula in which all of the variables are bound. We consider the standard model of number theory (as defined by the Peano axioms). A sentence is *provable* if it can be deduced from these axioms. The theory of arithmetic $L_a = (\mathbb{Z}_+, +, \cdot, =, <, 0, 1)$ is the collection of provable sentences. A relatively straightforward construction shows that L_c reduces to L_a, which yields the following fundamental result.

Theorem 1.3. *L_a is undecidable.*

This theorem implies the incompleteness of this model of number theory, i.e., there are sentences such that neither it nor its negation is provable. Every complete model is decidable, since a Turing machine can generate all possible deductions and stop if either the statement or its negation is proved.

These results were only the first steps in a rich area of research that can be viewed as the ancestor of modern-day complexity theory. One of its pinnacles of achievement is the solution of Hilbert's tenth problem, which asked for a procedure to decide if a given multivariate polynomial has an integer-valued root. Culminating years of progress in this area, Matijasevič proved that this problem is undecidable. [For an introduction to this history and the many results on which this proof builds, the reader is referred to the survey of Davis (1977).]

An important generalization of a Turing machine that will play a fundamental role in complexity theory is that of a *nondeterministic* Turing machine. Here, the transition function δ is no longer a function, but rather a relation, in that at each step there is a finite set of possible next moves of which exactly one is made. The notion of acceptance by a nondeterministic Turing machine is central to its definition: an input is accepted if there exists a sequence of transitions of δ that cause the machine to halt in an accepting state. A nondeterministic Turing machine M solves a decision problem L if L is the set of inputs accepted by M, and for every input M halts on each sequence of transitions. An equivalent formulation is to think of a nondeterministic Turing machine as a deterministic Turing machine with an additional *guess tape*, that is a read-only tape, where the head only moves to the right. The contents of the guess tape are magically constructed and presented to the machine as it begins the computation. For simplicity, we shall assume that the machine makes the same number of transitions for any guess. The set of inputs in L is the set of inputs for which there is a guess that allows the machine to halt in an accepting state. Note that a nondeterministic Turing machine accepting L can be converted into a deterministic machine for L by trying all of the (exponentially many) guesses. We can view a particular computation as proving (or disproving) the theorem "$x \in L$"; in this context, the difference between determinism and nondeterminism is analogous to the difference between proving a theorem and verifying its proof.

Consider again the hamiltonian circuit problem. A nondeterministic Turing machine for it might be constructed by letting the guess tape encode a sequence of n nodes of the graph. The Turing machine simply verifies that there is an arc between each consecutive pair of nodes in the guessed sequence, as well as between the first and last nodes. If the graph is hamiltonian, then there is a correct guess, but otherwise, for any sequence of n nodes there will be some pair that is not adjacent. The correct guess, in essence the hamiltonian circuit, is a *certificate* that the graph is hamiltonian. Observe that this definition is one-sided, since the requirements for instances in L and not in L are quite different.

1.3. Computational resources and complexity classes

Now that we have mathematical formulations of both problems and machines, we can describė what is meant by the computational resources required to solve a certain problem.

In considering the execution of a deterministic Turing machine, it is clear that the number of transitions before halting corresponds to the running time of the machine on a particular input. In discussing the running time of an algorithm, within any of the models, we do not want to simply speak of particular instances, and so we must make some characterization of the running time for all instances. The criterion that we will focus on for most of this chapter is the worst-case running time as a function of the input size. When we say that a Turing machine takes n^2 steps, this means that for any input of size n, the Turing machine always halts within n^2 transitions. Unless otherwise specified, we will let $n = |x|$ denote the length of the input string x.

We will not be interested in the precise count of the number of transitions, but rather in the order of the running time. A function $f(n)$ is $\mathrm{O}(g(n))$ if there are constants N and c such that for all $n \geqslant N$, $f(n) \leqslant cg(n)$. Thus, rather than say that a Turing machine has worst-case running time $3n^2 + 5n - 17$, we say simply that it is $\mathrm{O}(n^2)$. This simplification makes it possible to discuss complicated algorithms without being overwhelmed by details. Furthermore, for any Turing machine with superlinear running time and any constant c, there exists another Turing machine to solve the same problem that runs c times faster than the original, that can be constructed by using an expanded work-tape alphabet.

We will also use a notation analogous to $\mathrm{O}(\cdot)$ to indicate lower bounds. A function $f(n)$ is $\Omega(g(n))$ if there are constants N and c such that for all $n \geqslant N$, $f(n) \geqslant cg(n)$. A function $f(n)$ is $\Theta(g(n))$ if it is both $\mathrm{O}(g(n))$ and $\Omega(g(n))$.

For a nondeterministic Turing machine, we define the running time to be the number of transitions in any computation path generated by the input. (Recall that we added the restriction that all computation paths must have the same length.) For a circuit, we will want to characterize the number of operations performed as a measure of time, so that the relevant parameter is the *size* of the circuit, which is the order of the graph representing it. For a RAM, there are two standard ways in which to count the running time on a particular input. In each operation, a RAM can, for example, add two numbers of unbounded length. In the *unit-cost* model, this action takes one step, independent of the lengths of the numbers being added. In the *log-cost* model, this operation takes time proportional to the lengths of the numbers added. These measures can have radically different values. Take 17, and repeatedly square it k times. With each squaring, the number of bits in the binary representation essentially doubles. Thus, although we have taken only k steps in the unit-cost model, the time required according to the log-cost model is exponential in k. This may seem artificial, but this problem can occur, for example, in Gaussian elimination, if it is not implemented carefully. Technically, we will use the log-cost model, in order to ensure that the RAM is equivalent to the Turing machine in the desired ways. But, when speaking of the running time of an algorithm, it is traditional to state the running time in the unit-cost model, since for all standard algorithms one can prove that the pathological behavior of the above example does not come into play.

Time is not the only computational resource in which we will be interested; we will see that the space complexity of certain combinatorial problems gives more insight into the structure of these problems. In considering the space requirements, we will focus on the Turing machine model, and will only count the number of cells used on the work tapes of the machine. Furthermore, we will again be interested in the asymptotic worst-case analysis of the space used. As was true for time bounds, the space used by a Turing machine can be compressed by any constant factor. The space used by a nondeterministic Turing machine is the maximum space used on any computation path. For circuits, it will also be interesting to study their *depth*, the longest path from a node of indegree 0 to one of outdegree 0.

The notion of the *complexity* of a problem is the order of a given computational resource, such as time, that is necessary and sufficient to solve the problem. Consider the following *directed reachability problem*: given a directed graph G, and two specified nodes s and t, does there exist a path from s to t? When we say that the complexity of the directed reachability problem for a graph with m arcs is $\Theta(m)$, this means that there is a (unit-cost RAM) algorithm that has worst-case running time $O(m)$ and there is no algorithm with running time of lower order. Tight results of this kind are extremely rare, since the tremendous progress in the design of efficient algorithms has not been matched, or even approached, by the slow progress in techniques for proving lower bounds on the complexity of these problems in general models of computation. For example, consider the 3-*colorability problem*: given an undirected graph G with m edges, can the nodes be colored with three colors so that no two adjacent nodes are given the same color; i.e., is $\chi(G) \leqslant 3$? The best lower bound is only $\Omega(m)$, in spite of substantial evidence that it cannot be solved in time bounded by a polynomial.

In order to study the relative power of particular computational resources, we introduce the notion of a *complexity class*, which is the set of problems that have a specified upper bound on their complexity. It will be convenient to define the complexity class DTIME($T(n)$) to be the set of all languages L that can be recognized by a deterministic Turing machine within time $O(T(n))$. NTIME($T(n)$)

denotes the analogous class of languages for nondeterministic Turing machines. Throughout this chapter, it will be convenient to make certain assumptions about the sorts of time bounds that define complexity classes. A function $T(n)$ is called *fully time-constructible* if there exists a Turing machine that halts after exactly $T(n)$ steps on any input of length n. All common time bounds, such as $n \log n$ or n^2, are fully time-constructible. We will *implicitly* assume that any function $T(n)$ used to define a time-complexity class is fully time-constructible.

The single most important complexity class is $\mathcal{P}$, the class of decision problems solvable in polynomial time. Two of the best-known algorithms, the Euclidean algorithm for finding the greatest common divisor of two integers and Gaussian elimination for solving a system of linear equations, are classical examples of polynomial-time algorithms. In fact, Lamé observed as early as 1844 that the Euclidean algorithm was a polynomial-time algorithm. In 1953, von Neumann contrasted the running time for an algorithm for the assignment problem that "turn[ed] out [to be] a moderate power of n, i.e., considerably smaller than the 'obvious' estimate $n!$" for a complete enumeration of the solutions. Edmonds (1965) and Cobham (1965) were the first to introduce $\mathcal{P}$ as an important complexity class, and it was through the pioneering work of Edmonds that polynomial solvability became recognized as a theoretical model of efficiency. With only a few exceptions, the discovery of a polynomial-time algorithm has proved to be an important first step in the direction of finding truly efficient algorithms. Polynomial time has proved to be very fruitful as a theoretical model of efficiency both in yielding a deep and interesting theory of algorithms and in designing efficient algorithms.

There has been substantial work over the last 25 years in finding polynomial-time algorithms for combinatorial problems. It is a testament to the importance of this development that much of this Handbook is devoted to discussing these algorithms. This work includes algorithms for graph connectivity and network flow (see chapter 2), for graph matchings (see chapter 3), for matroid problems (see chapter 11), for point lattice problems (see chapter 19), for testing isomorphism (see chapter 27), for finding disjoint paths in graphs (see chapter 5) as well as for problems connected with linear programming (see chapters 28 and 30).

Another, more technical reason for the acceptance of $\mathcal{P}$ as the theoretical notion of efficiency, is its mathematical robustness. Recall the discussion of encodings where we remarked that any reasonable encoding will have length bounded by a polynomial in the length of another. As a result, any polynomial-time algorithm which expects its input in one form can be converted to a polynomial-time algorithm for the other. In particular, note that the previous discussion of the two different encodings of a graph can be swept aside and we can assume that the size of the input for a graph of order n is n. Notice further that the informal definition of $\mathcal{P}$ given above does not rely on any model of (deterministic) computation. One justification for this statement is the following theorem.

Theorem 1.4. *The following classes of problems are identical:*

- *the class of computational problems solvable by a Turing machine in polynomial-time;*

- *the class of computational problems solvable by a RAM in polynomial-time under the log-cost measure;*
- *the class of computational problems solvable by a family of circuits of polynomial size, where the circuit for inputs of size n can be generated by a Turing machine with running time bounded by a polynomial in n.*

The importance of the class $\mathcal{NP} = \bigcup_k \mathrm{NTIME}(n^k)$ is due to the wide range of important problems that are known to be in the class, and yet are not known to lie in $\mathcal{P}$. For example, the nondeterministic algorithm given for the hamiltonian circuit problem is clearly polynomial, and so this problem lies in $\mathcal{NP}$. However, it is not known whether this or any problem is in $\mathcal{NP} \setminus \mathcal{P}$, and this is, undoubtably the central question in complexity theory.

Open Problem. Is $\mathcal{P} = \mathcal{NP}$?

The following reformulation of $\mathcal{NP}$ is often useful: $L \in \mathcal{NP}$ if there exists a language $L' \in \mathcal{P}$ and a polynomial $p(n)$ such that $x \in L \Leftrightarrow \exists y$ such that $|y| = p(|x|)$ and $(x, y) \in L'$. (We will denote this polynomially bounded quantification by $\exists_p$.)

For each decision problem L, there is a complementary problem, $\bar{L}$, such as the problem of recognizing non-hamiltonian graphs. For any complexity class $\mathcal{S}$, let co-$\mathcal{S}$ denote the class of languages whose complement is in $\mathcal{S}$. The definition of $\mathcal{P}$ is symmetric with respect to membership and non-membership in L, so that $\mathcal{P} = \text{co-}\mathcal{P}$. In this respect $\mathcal{NP}$ is very different. In fact, it is unknown whether the hamiltonian circuit problem is in co-$\mathcal{NP}$.

Open Problem. Is $\mathcal{NP} = \text{co-}\mathcal{NP}$?

Edmonds (1965) brought attention to the class $\mathcal{NP} \cap \text{co-}\mathcal{NP}$, and called problems in this class *well-characterized*, since there is a short certificate to show that the property holds, as well as a short certificate that it does not. Edmonds was working on algorithms for non-bipartite maximum matching at the time, and this problem serves as a good example of a problem in this class. If the instance consists of a graph G and a bound k and we wish to know if there is a matching of size at least k, the matching itself serves as a certificate for an instance in L, whereas an odd-set cover serves as a certificate for an instance not in L (see chapter 3). Note that there is a min–max theorem characterizing the size of the maximum matching that is at the core of the fact that matching is in $\mathcal{NP} \cap \text{co-}\mathcal{NP}$, and indeed min–max theorems often serve this role. As mentioned above, matching is known to be in $\mathcal{P}$, and this raises the following question.

Open Problem. Is $\mathcal{P} = \mathcal{NP} \cap \text{co-}\mathcal{NP}$?

We will also be concerned with complexity classes defined by the space complexity of problems. As for time, let DSPACE($S(n)$) and NSPACE($S(n)$) denote, respectively, the class of languages accepted by deterministic and nondeterministic Turing machines within space O($S(n)$). We will implicitly assume the following condition for all space bounds used to define complexity classes: a function $S(n)$ is *fully space-constructible* if there is a Turing machine that, on any input of length n delimits $S(n)$ tape cells and then halts. Three space complexity classes will receive the most prominent attention:

- $\mathcal{L} = \mathrm{DSPACE}(\log n)$;
- $\mathcal{NL} = \mathrm{NSPACE}(\log n)$; and
- $\mathcal{PSPACE} = \bigcup_k \mathrm{DSPACE}(n^k)$.

One might be tempted to add a fourth class, $\mathcal{NPSPACE}$, but we shall see that nondeterminism does not add anything in this case. We will see that the chain of inclusions

$$\mathcal{L} \subseteq \mathcal{NL} \subseteq \mathcal{P} \subseteq \mathcal{NP} \subseteq \mathcal{PSPACE}$$

holds, and the main thrust of complexity theory is to understand which of these inclusions is proper. At the extremes, a straightforward diagonalization argument due to Hartmanis, Lewis and Stearns shows that $\mathcal{L} \neq \mathcal{PSPACE}$, and a result of Savitch further implies that $\mathcal{NL} \neq \mathcal{PSPACE}$, but after nearly a quarter century's more work, these are the only sets in this chain known to be distinct.

1.4. Randomized computation

In this subsection, we will consider models of computation that exploit the power of randomization in the design and analysis of algorithms *without* making probabilistic assumptions about the inputs. A randomized algorithm is an algorithm that can flip coins during the computation; that is, we consider a fixed input, and study the algorithm's behavior as a random variable depending only on the coin flips used. For the algorithms discussed below, the algorithm is allowed to make mistakes, but for each input the probability of error must be very small.

The best-known randomized algorithm is for testing primality. In the *primality testing problem*, we are given a natural number N, and we wish to decide if it is prime. It is not known whether primality testing is in $\mathcal{P}$. The input size of the number N is $\log N$, and therefore algorithms that simplistically search for divisors of N do not run in polynomial time. Consider Fermat's theorem: if N is prime then a^N is congruent to a modulo N for every integer a. This provides a way to conclude that a number is not prime without actually exhibiting a factor. That is, if we find an integer a such that a^N is not congruent to a modulo N, denoted $a^N \not\equiv a \bmod N$, then we can conclude that N is not prime. (Note that $a^N \bmod N$ can be computed in polynomial time by repeatedly squaring modulo N.) Let such an a be called a *witness* for N's compositeness. The advantage of this kind of witness, compared to exhibiting a divisor, is that if there exists an integer a such that $a^N \not\equiv a \bmod N$, then at least half of the integers in the range from 1 to N have this property.

Unfortunately, there are composite numbers, the so-called *Carmichael numbers* that are not prime, but for which no witness exists. If we momentarily forget about the existence of these numbers, we get the following algorithm for testing primality: given an integer N, choose an integer a in the range 1 to N at random, and check if a is a witness for N. If a witness is found, then we know that N is not prime (and not even a Carmichael number). On the other hand, if N is not a prime (and also not a Carmichael number), then a random a is a witness with probability at least one half. Running this test k times with independent random choices, we either find a witness or can fairly safely conclude that no witness exists (with error probability 2^{-k}). This gives a randomized polynomial-time algorithm to recognize the language of all primes and Carmichael numbers. Rabin (1976) and separately Solovay and Strassen (1977), by using a somewhat more sophisticated variant of Fermat's theorem, gave randomized polynomial-time algorithms that accept the language of all primes.

The above idea actually gives a polynomial-time algorithm if the extended Riemann hypothesis is true. Miller has proved that if the extended Riemann hypothesis holds, then there exists a witness for N (in more or less the above sense) that is at most $O((\log N)^2)$. By trying all the integers up to this limit, we would get a polynomial-time deterministic algorithm for primality testing. For more details on this and other number-theoretic algorithms see the survey of Lenstra and Lenstra (1990).

The formal definition of a *randomized Turing machine* is similar to the definition of a nondeterministic Turing machine in the sense that at every point during the computation there could be several different next steps. Randomized Turing machines have a read-only *randomizing tape* similar to the guess tape of the nondeterministic Turing machine. We can think of this tape as providing the outcomes of the coin flips to be used by the algorithm. We shall assume that for a given input length n, the algorithm reads a fixed number of bits, $f(n)$, from the randomizing tape. The *probability* that the randomized Turing machine accepts an input x of length n is defined to be the fraction of all possible strings of length $f(n)$ that, when used as the initial segment of the randomizing tape, cause the Turing machine to accept x. For the randomized Turing machine, we take the simplifying approach that the running time for an input is the maximum number of transitions in some sequence of allowed transitions (i.e., for some contents of the randomizing tape). Unlike the nondeterministic Turing machine, a randomized Turing machine is not just a convenient mathematical model; randomized algorithms can be implemented in practical settings.

We define $\mathcal{BPP}$, the class of languages accepted by a randomized polynomial-time algorithm, as follows. A language L is in $\mathcal{BPP}$ if there exists a polynomial-time randomized Turing machine that accepts each $x \in L$ with probability at least $\frac{2}{3}$; and rejects each $x \notin L$ with probability at least $\frac{2}{3}$. We can think of the outcomes of the computation as follows: if the Turing machine accepts x this means that "x is probably in L", whereas if it rejects, that means that "x is probably not in L". Note that the choice of the number $\frac{2}{3}$ in the definition was rather arbitrary: if we run the algorithm k times independently, and take the majority decision, we can decrease the probability of error exponentially in k. If k is fairly large, then one can accept the answer given by the algorithm without any reasonable shadow of doubt. (The letters $\mathcal{BPP}$ stand for a Probabilistic Polynomial-time algorithm with probabilities Bounded away from $\frac{1}{2}$.)

Other problems not known to be in $\mathcal{P}$ for which there is a randomized polynomial-time algorithm include computing the square root of an integer x modulo a prime p, and deciding whether the determinant of a matrix whose entries are multivariate polynomials is the zero polynomial. The algorithm for the latter problem assigns random values to the variables in the polynomials, and computes the resulting determinant (of numbers). If this determinant is non-zero, then certainly the determinant of variables is non-zero as a polynomial. On the other hand, if the determinant of variables is non-zero, then a random evaluation will yield a non-zero determinant with very high probability. This can be used to show that given a graph G, a subset of edges F and an integer k, determining whether a graph has a perfect matching with exactly k edges in F can be solved by a randomized polynomial-time algorithm (see chapter 3).

Notice that the randomized primality testing algorithm has a property stronger than required by the formal definition. The conclusion that N is not a prime was certain; uncertainty arose only in the case of the conclusion "N is probably prime". The complexity class $\mathcal{RP}$ is defined to reflect this asymmetry. A language L is in $\mathcal{RP}$ if there exists a polynomial-time randomized Turing machine RM such that each input that RM can accept (along any computation path) is in L and for each input $x \in L$, the probability that RM accepts x is at least $\frac{1}{2}$. Note again that the choice of the number $\frac{1}{2}$ is arbitrary.

The above mentioned algorithms show that primality testing is in co-$\mathcal{RP}$. There are very few problems known to be in $\mathcal{BPP}$ but not in $\mathcal{RP}$ or co-$\mathcal{RP}$. Bach, Miller and Shallit provided the first "natural" examples of problems in $\mathcal{BPP}$ that are not obviously in $\mathcal{RP}$ or co-$\mathcal{RP}$. They proved that the set of perfect numbers is in $\mathcal{BPP}$. (A natural number N is *perfect* if the sum of all its natural divisors is $2N$; for example, 6 is perfect.)

In some sense, an $\mathcal{RP}$ algorithm is more satisfying than a $\mathcal{BPP}$ algorithm, since at least one of the two conclusions reached can be claimed with certainty. A randomized algorithm that *never makes mistakes* would be even more desirable. This can be defined in the following way: an algorithm to compute a function $f(x)$ is a *Las Vegas algorithm* if, given an input x, this randomized algorithm either correctly computes $f(x)$, or it halts without coming to a conclusion, and the probability of the latter outcome is less than $\frac{1}{2}$ for each input x. It is easy to see that a language L is in $\mathcal{RP} \cap$ co-$\mathcal{RP}$ if and only if there exists a polynomial-time Las Vegas algorithm to decide membership in L. Observe also that if we repeat any Las Vegas algorithm until it gives an answer, the resulting algorithm always gives the correct answer, and for any input, it is expected to run in polynomial time. Extending results of Goldwasser and Kilian, Adleman and Huang give a sophisticated Las Vegas algorithm for testing primality.

There is evidence that suggests that randomized polynomial time is not that different from $\mathcal{P}$. For example, consider a nonuniform analog of $\mathcal{P}$ by considering the class of languages accepted by a family of polynomial-size circuits. Alternatively, one can make the Turing machine model nonuniform, by allowing the machine free access to a prespecified polynomial-length advice string s_n, when processing any input of length n. This class is denoted $\mathcal{P}$/poly. For any language L in $\mathcal{BPP}$, we can assume without loss of generality that there is a machine RM that accepts each $x \in L$ with probability $1-2^{-(n+1)}$ and rejects each $x \notin L$ with probability $1-2^{-(n+1)}$, where n denotes the length of x (since taking the majority decision of repeated trials decreases the error probability exponentially). However, this implies that the probability that a random string used by RM for an input of length n would work correctly for *all* strings of length n is at least $\frac{1}{2}$. Thus, there must exist such a good string to serve as the advice, and we have shown the following theorem, which is based on an idea of Adleman.

Theorem 1.5. $\mathcal{BPP} \subseteq \mathcal{P}$/poly.

2. Shades of intractability

In this section we will consider many computational problems, and see that the universe does not appear to be divided simply into tractable and intractable problems. Current evidence suggests that there are a variety of different classes of problems, each characterizing its own particular shade of intractability. Much of the work in complexity theory is aimed at understanding the correct framework in which to place these problems.

The subsections here reflect three types of approaches for characterizing the difficulty of these problems. The nicest sort of result places absolute limits on our ability to solve problems; for example, the most severe limit is to show that a problem is undecidable, and among decidable problems there are only a handful that can be proven intractable, in the sense that they require a certain (non-trivial) amount of time or space to be solved. Much more common is to provide a completeness result to show that a particular problem is a hardest problem within a given complexity class. If the class contains a great number of problems not known to be solvable with more modest resources, this provides evidence that the problem is intractable. Finally, in order to better understand a problem, it has frequently been useful to strengthen the basic Turing machine model in order define complexity classes that better characterize the problem. Such an alternative view has often made problems appear less intractable; the subsections on the polynomial-time hierarchy and on randomized proofs present results in this direction.

2.1. *Evidence of intractability: $\mathcal{NP}$-completeness*

The lack of lower bounds with respect to a general model of computation for either space or time complexity has led to the search for other evidence that suggests that lower bounds hold. One such type of evidence might be the implication: if the hamiltonian circuit problem is in $\mathcal{P}$, then $\mathcal{P}=\mathcal{NP}$. Now, $\mathcal{NP}$ contains a tremendous variety of problems that are not known to be in $\mathcal{P}$, and so by proving such a claim, one shows that the hamiltonian circuit problem is a hardest problem in $\mathcal{NP}$, in that any polynomial-time algorithm to solve it would in fact solve thousands of other problems.

The principal tool in providing evidence of this form is that of reduction. We will say that a problem L_1 *polynomial-time reduces* to L_2 if there exists a polynomial-time computable function f that maps instances of L_1 into instances of L_2 such that x is a "yes" instance of L_1 if and only if $f(x)$ is a "yes" of L_2. We shall denote this by $L_1 \propto_p L_2$. Notice that if there were a polynomial-time algorithm for L_2, we could then obtain a polynomial-time algorithm for L_1 by first computing $f(x)$ and then running the assumed algorithm on $f(x)$. This composite procedure is polynomial-time, since the composition of two polynomials is itself a polynomial.

Definition. A problem L_1 is *$\mathcal{NP}$-complete* if
(i) $L_1 \in \mathcal{NP}$;
(ii) for all $L \in \mathcal{NP}$, $L \propto_p L_1$.

The composition argument given above yields the following results.

Theorem 2.1. *For any $\mathcal{NP}$-complete problem L, $L \in \mathcal{P}$ if and only if $\mathcal{P} = \mathcal{NP}$.*

Theorem 2.2. *If L_1 is $\mathcal{NP}$-complete, $L_2 \in \mathcal{NP}$ and $L_1 \propto_p L_2$, then L_2 is $\mathcal{NP}$-complete.*

The first result says that any $\mathcal{NP}$-complete problem completely characterizes $\mathcal{NP}$ in its relationship to $\mathcal{P}$, and that we can focus on any such problem without loss of generality in trying to prove that the two classes are different. The hamiltonian circuit problem is $\mathcal{NP}$-complete, and in this section we will prove that several combinatorial problems are among the plethora having this property. The second result gives a strategy for proving that a problem is $\mathcal{NP}$-complete, provided that a "first" $\mathcal{NP}$-complete problem is known. Of course, to initiate this strategy, we must show that some natural problem has this property, and it was a landmark achievement in complexity theory when Cook (1971) showed that the problem of deciding the satisfiability of a formula in propositional logic is $\mathcal{NP}$-complete. The importance of this result was only fully recognized through the work of Karp (1972), whose seminal paper showed 21 well-known combinatorial problems to be $\mathcal{NP}$-complete. Independently, Levin (1973) discovered the same approach to studying the intractability of computational problems.

A Boolean formula in conjunctive normal form is the conjunction (and) of clauses $C_1, \ldots, C_s$, each of which is a disjunction (or) of literals $x_1, \bar{x}_1, \ldots, x_t, \bar{x}_t$, where each x_i is a Boolean variable and $\bar{x}_i$ denotes its negation. In the *satisfiability problem* (SAT) we are given such a Boolean formula, and asked to decide if there exists a truth assignment for the variables such that the formula evaluates to true.

Theorem 2.3. *The satisfiability problem is $\mathcal{NP}$-complete.*

Proof. It is easy to see that the satisfiability problem is in $\mathcal{NP}$, since we can interpret the first t cells of the guess tape as providing the assignment, and then it is a simple matter to evaluate the formula for that assignment in polynomial time.

Next, we must show that for any language $L \in \mathcal{NP}$ there exists a polynomial-time function f that maps each instance x of the original problem into a Boolean formula such that $x \in L$ if and only if $f(x)$ is satisfiable. Before giving the reduction, we first argue that M may be assumed to have a form somewhat simpler than the original definition. Imagine that the Turing machine has only one tape, which serves both as the input tape and as the work tapes. A simple construction shows that if there exists a nondeterministic Turing machine M with running time $T(n)$, then there exists a nondeterministic machine of this simpler form that finishes within time $T^2(n)$ (by enlarging the alphabet so that each symbol denotes one character on all of the tapes of M). Furthermore, we can assume without loss of generality that for some l the machine M runs for exactly time n^l on every input of length n.

Let M be such a simplified machine that runs on input x for T steps before halting. We can describe the configuration of the machine at any instant of the computation by giving the contents of the tape, the position of the head and the current state. This can all be encoded as a string by using the alphabet $\Gamma' = \Gamma \times (Q \cup \{!\})$ where the first coordinate gives the contents of a cell of the tape, and the second is '!' unless the head is reading that cell of the tape, when it is the current state. We can then encode the entire computation as a matrix, where each row of the matrix corresponds to one step of the computation, and each column corresponds to a cell of the tape. (Note that there are no more than T cells in either direction of the initial head position that could be reached during the computation.) Acceptance of x by M boils down to the following question: does there exist a guess that causes the matrix to be filled in so that in the last row, the configuration contains an accepting state?

It is straightforward to construct a Boolean formula that represents this question. Let $g_1, \ldots, g_T$ be variables that represent the binary values of the guess tape. Let a_{ijk} represent the contents of the ijth cell of the matrix in the sense that it is 1 if and only if it contains the kth character of Γ'. The formula will be a conjunction of pieces that correspond to the following conditions that we wish to impose: the variables represent some matrix, in that exactly one character is stored in each entry; the first row corresponds to the initial configuration for input x; the last row contains an accepting state; the computation proceeds in the deterministic way specified by the guesses g_i. We will not give each of these in detail, but only sketch the main ideas. The first is easy: for each of the $O(T^2)$ entries, check that at least one of the associated variables is 1 (by their or) and for each pair, check that not both are 1 (by the or of their negations). The second and third are equally routine. The last condition takes a bit more work and is based on a principle of locality: if locally the computation (i.e., the matrix) appears correct, then the entire computation was performed correctly. In fact, it is sufficient to check that each 2×2 submatrix appears correct. Furthermore, it is an easy exercise to encode that some 2×2 submatrix in the ith and $(i+1)$th rows behaves according to the guess g_i. By taking the conjunction of all $O(T^2)$ such pieces of local information, we enforce that the computation is done correctly. It is now routine to verify that the formula constructed is satisfiable if and only if there are guesses that lead M to accept x. □

Now that we have our initial $\mathcal{NP}$-complete problem, we proceed to give a number of reductions to show that several important combinatorial problems are $\mathcal{NP}$-complete. Literally thousands of problems are now known to be $\mathcal{NP}$-complete, so we will only present a small handful of examples that serve to illustrate an important phenomenon in complexity theory, or relate to important combinatorial problems discussed elsewhere in this chapter, as well as in the rest of this volume. Most of the problems that we consider were shown to be $\mathcal{NP}$-complete in the pioneering work of Karp (1972).

Many restricted cases of the satisfiability problem are also $\mathcal{NP}$-complete. One that is often used in further $\mathcal{NP}$-completeness proofs is the 3-SAT problem, where each clause of the conjunctive normal form must contain exactly three literals. It is a simple task to show that by adding additional variables, longer clauses can be broken into clauses of length three, yielding a new formula that can be satisfied if and only if the original can.

For the *stable set problem*, we are given a graph G and a bound k, and asked to decide if $\alpha(G) \geqslant k$; that is, do there exist k pairwise non-adjacent nodes in G? It is easy to see that this problem is in $\mathcal{NP}$, and to show that it is complete, we reduce 3-SAT to it and invoke Theorem 2.2. Given a 3-SAT instance ϕ, we construct a graph G as follows: for each clause in ϕ, let there be three nodes in G, each representing a literal in the clause, and let these three nodes induce a *clique* (i.e., they are pairwise adjacent); complete the construction by making adjacent any pair of nodes that represent a literal and its negation, and set k to be the number of clauses in ϕ. If there is a satisfying assignment for ϕ, pick one literal from each clause that is given the assignment true; the corresponding nodes in G form a stable set of size k. If there is a stable set of size k, then it must have exactly one node in each clique corresponding to a clause. Furthermore, the nodes in the stable set cannot correspond to both a literal and its negation, so that we can form an assignment by setting to true all literals selected in the stable set, and extending this assignment to the remaining variables arbitrarily. This is a satisfying assignment. This is a characteristic reduction, in that we build *gadgets* to represent the variables and clause structure within the framework of the new problem. The $\mathcal{NP}$-completeness of two other problems follow immediately: the *clique problem*, given a graph G and a bound k, decide if there is a clique in G of size k; and the *node cover problem*, given a graph G and a bound k, decide if there exists a set of k nodes such that every edge is incident to a node in the set. A somewhat more complicated reduction transforms 3-SAT into the hamiltonian circuit problem to show that to be $\mathcal{NP}$-complete. A seemingly slight generalization of bipartite graph matching, the 3-*dimensional matching problem*, can be shown to be $\mathcal{NP}$-complete: given disjoint node sets A, B and C, and a collection $\mathcal{F}$ of hyperedges of the form (a,b,c) where $a \in A$, $b \in B$ and $c \in C$, does there exist a subset of $\mathcal{F}$ such that each node is contained in exactly one edge in the subset?

If we restrict the stable set problem to a particular constant value of k (e.g., is $\alpha(G) \leqslant 100$?), then this problem can be solved in $\mathcal{P}$ by enumerating all possible sets of size 100. In contrast to this, Stockmeyer has shown that the 3-colorability problem is $\mathcal{NP}$-complete, by reducing 3-SAT to it. Let ϕ be a 3-SAT formula. We construct a graph G from it in the following way. First, construct a "reference"

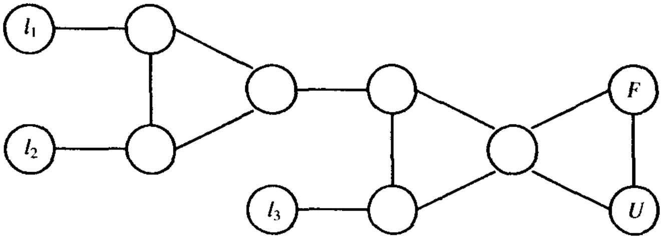


Figure 1.

clique on three nodes, called true, false and undefined; these nodes will serve as a way of naming the colors in any 3-coloring of the graph. For each variable in ϕ, construct a pair of adjacent nodes, one representing the variable, and one representing its negation, and make them both adjacent to the undefined node. For each clause, $l_1 \vee l_2 \vee l_3$, construct the subgraph shown in fig. 1, where the nodes labeled with literals, as well as false (F) and undefined (U), are the nodes already described. It is easy to see that if ϕ has a satisfying assignment, we can get a proper 3-coloring of this graph as follows: color the nodes corresponding to literals that are true in the assignment with the same color as is given node true; color analogously the nodes for false literals; and then extend this coloring to the remaining nodes in a straightforward manner. Furthermore, it involves only a little case-checking to see that if the graph is 3-colorable, then the colors can be interpreted as a satisfying assignment.

The *integer programming problem*, defined as follows, is $\mathcal{NP}$-complete: given an $m \times n$ matrix A and an m-vector b, decide if there exists an integer n-vector x such that $Ax \leqslant b$. In this case, finding a reduction from 3-SAT is trivial: given a formula ϕ, represent each literal by an integer variable bounded between 0 and 1, and for each Boolean variable x, constrain the sum of the variables corresponding to x and its negation to be at most 1. The construction is completed by adding a constraint for each clause that forces the variables for the literals in the clause to sum to at least 1. On the other hand, to show that the problem is in $\mathcal{NP}$ requires more work, involving a calculation that bounds the length of a "smallest" solution satisfying the constraints (if one exists at all).

The *subset sum problem* is also $\mathcal{NP}$-complete: given a set of numbers S and a target number t, does there exist a subset of S that sums to t? This is the first problem that we have encountered that is a "number problem" in the sense that it is not the combinatorial structure, but rather the numbers that make this problem hard. If the numbers are given in unary, there is a polynomial-time algorithm (such an algorithm is called *pseudo-polynomial*): keep a (large, but polynomially bounded) table of all possible sums obtainable using a subset of the first i numbers; this is trivial to do for $i = 1$, and it is not hard to efficiently find the table for $i+1$ given the table for i; the table for $i = n$ gives us the answer. There are "number problems" that remain $\mathcal{NP}$-complete, even if the numbers are encoded in unary; such problems are called *strongly $\mathcal{NP}$-complete*. One example is the *3-partition problem*: given a set S of the $3n$ integers that sum to nB, does there exist a partition of S into n sets, $T_1, \ldots, T_n$, where each T_i has three elements that sum to B?

2.2. The complexity class $\mathcal{NP} \setminus \mathcal{P}$?

If we are to believe the conjecture that $\mathcal{P} \neq \mathcal{NP}$, then there exists a non-empty complexity class $\mathcal{NP} \setminus \mathcal{P}$. One might ask the question: is it true that every problem in $\mathcal{NP}$ is either $\mathcal{NP}$-complete or in $\mathcal{P}$? If $\mathcal{P} = \mathcal{NP}$ this question has a (trivial) affirmative answer, but a negative answer to it (under the assumption that $\mathcal{P} \neq \mathcal{NP}$) might help explain the reluctance of certain problems to be placed in one of those two classes. In fact, Ladner has shown, under the assumption that $\mathcal{P} \neq \mathcal{NP}$, that there is an extremely refined structure of equivalence between the two classes, $\mathcal{P}$ and $\mathcal{NP}$-complete.

Theorem 2.4. *If L_1 is decidable and $L_1 \notin \mathcal{P}$, then there exists a decidable language L_2 such that $L_2 \notin \mathcal{P}$, $L_2 \propto_{\mathrm{p}} L_1$ but $L_1 \not\propto_{\mathrm{p}} L_2$.*

Note that if L_1 is $\mathcal{NP}$-complete, the language L_2 is "in between" the classes $\mathcal{P}$ and $\mathcal{NP}$-complete, if $\mathcal{P} \neq \mathcal{NP}$, and by repeatedly applying the result we see that there is a whole range of seemingly different complexity classes. Under the assumption that $\mathcal{P}$ is different from $\mathcal{NP}$, do we know of any candidate problems that may lie in this purgatory of complexity classes? The answer to this is "maybe". We will give four important problems that have not been shown to be either in $\mathcal{P}$ or $\mathcal{NP}$-complete. The problem for which there has been the greatest speculation along these lines is the *graph isomorphism problem*: given a pair of graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$, decide if there is a bijection $\sigma : V_1 \mapsto V_2$ such that $ij \in E_1$ if and only if $\sigma(i)\sigma(j) \in E_2$. Later in this section we will provide evidence that it is *not* $\mathcal{NP}$-complete, and efforts to show that it is in $\mathcal{P}$ have so far fallen short (see chapter 27 of this volume). A problem that mathematicians since the ancient Greeks have been trying to solve is that of *factoring* integers; a decision formulation that is polynomially equivalent to factoring is as follows: given an integer N and a bound k, does there exist a factor of N that is at most k? A problem that is no harder than factoring is the *discrete logarithm problem*: given a prime p, a generator g and a natural number $x < p$, find l such that $g^l \equiv x \bmod p$. Finally, there is the *shortest vector problem*, where we are given a collection of integer vectors, and we wish to find the shortest vector (in the Euclidean norm) that can be represented as a non-zero integral combination of these vectors. Here, current evidence makes it seem likely that this problem is really $\mathcal{NP}$-complete; related work is discussed elsewhere in this volume (see chapter 19).

It is important to mention that there is an important subclass of $\mathcal{NP}$ which may also fall in this presumed gap. Edmonds' class of well-characterized problems,

$\mathcal{NP} \cap$ co-$\mathcal{NP}$, certainly contains $\mathcal{P}$ and is contained in $\mathcal{NP}$. Furthermore, unless $\mathcal{NP}$ = co-$\mathcal{NP}$, it cannot contain any $\mathcal{NP}$-complete problem. On the other hand, the prevailing feeling is that showing a problem to be in this class is a giant step towards showing that the problem is in $\mathcal{P}$. A result of Pratt shows that primality is in $\mathcal{NP}$, so it lies in $\mathcal{NP} \cap$ co-$\mathcal{NP}$, though as discussed above, there is additional evidence that it lies in $\mathcal{P}$. The factoring problem, which appears to be significantly harder, also lies in $\mathcal{NP} \cap$ co-$\mathcal{NP}$, since a prime factorization can be guessed along with a certificate that each of the factors is indeed prime. One interesting open question connected with $\mathcal{NP} \cap$ co-$\mathcal{NP}$ is concerned with the existence of a problem that is complete for this class. One might hope that there is some natural problem that completely characterizes the relationship of $\mathcal{NP} \cap$ co-$\mathcal{NP}$ with $\mathcal{P}$ (in the same manner that 3-SAT characterizes the $\mathcal{P} = \mathcal{NP}$ question).

One approach to shedding light on the complexity of a problem that is not known to be either in $\mathcal{P}$ or $\mathcal{NP}$-complete, has been to consider weaker forms of completeness for $\mathcal{NP}$. In fact, Cook's notion of completeness, though technically a weaker definition of intractability, is no less damning. $L_1 \propto_{\mathrm{p}} L_2$ can be thought of as solving L_1 by a restricted kind of subroutine call, where first some polynomial-time preprocessing is done, and then the subroutine for L_2 is called once. Cook (1971) proposed a notion of reducibility where L_1 is solved by using a polynomial-time Turing machine that can, in one step, get an answer to any query of the form "is $x \in L_2$?". Note that any complete language L with respect to this reducibility still has the property that $L \in \mathcal{P}$ if and only if $\mathcal{P} = \mathcal{NP}$. Karp (1972) focused attention on the notion $\propto_{\mathrm{p}}$, and was able to show that $\mathcal{NP}$-completeness was powerful enough to capture a wide range of combinatorial problems. On the other hand, it remains an open question to show that Cook's notion of reducibility is stronger than Karp's; is there a natural problem in $\mathcal{NP}$ that is complete with respect to "Cook" reducibility, but not with respect to "Karp" reducibility? Adleman and Manders showed that non-determinism and randomization can play a role in defining notions of reducibility, and used these notions to show that certain number theoretic problems were not, for example, in $\mathcal{NP} \cap$ co-$\mathcal{NP}$ unless $\mathcal{NP}$ = co-$\mathcal{NP}$.

2.3. *Oracles and relativized complexity classes*

In previous subsections, we have discussed techniques to provide evidence of the intractability of concrete problems in $\mathcal{NP}$ by proving completeness results. In this section, we will be concerned with extensions of the model of computation where the analog of the $\mathcal{P}$ versus $\mathcal{NP}$ problem can be resolved.

We have mentioned earlier that oracles are used to compactly represent the input of some problems. One can define the analog of the classes $\mathcal{P}$ and $\mathcal{NP}$ for problems whose inputs are oracles and it is easy to prove that these complexity classes are different. For example, if the oracle represents a set system on n elements, then the problem to decide if this set system is nonempty is clearly in $\mathcal{NP}$, but one has to ask 2^n queries of the oracle to resolve it deterministically. Similar results have also been proved for combinatorial problems that are naturally represented by oracles. For example, the matroid matching problem and the problem to decide if a matroid has girth at most k (i.e., whether it has a cycle of length at most k)

are both clearly in $\mathcal{NP}$, but it has been shown that any deterministic algorithm solving them has to ask an exponential number of queries (see chapter 11).

Problems on graphs can also be given by an oracle. Suppose that a graph (e.g., on $N = 2^n$ nodes) is given by an oracle that can tell whether two nodes are adjacent. The question is whether all "reasonable" decision problems on graphs require one to ask some constant fraction of the queries? This problem has a long history, both for directed and undirected graphs, and many attempts were made at giving sufficiently strong conditions before an accurate conjecture, due to Aanderaa and Rosenberg, was proved by Rivest and Vuillemin, and later strengthened by Kahn, Saks and Sturtevant. Consider a decision problem L where the instances are undirected graphs, and L has three important properties: (1) *nontriviality* – some graphs are in L, but not all; (2) *monotonicity* – if G is an instance in L and G' is formed by adding an edge, then G' is also in L ; (3) *invariance under isomorphism* – if G is an instance in L and its nodes are relabeled to form G', then G' is also in L. Then for any problem satisfying these two properties, any correct procedure uses essentially $N^2/4$ queries for some graph of order N. In fact, if N is a prime power, Kahn, Saks and Sturtevant have shown that all $N(N-1)/2$ queries must be asked (see chapter 34). These results are also relevant in comparing the adjacency-matrix form of encoding graphs to the adjacency-list encoding.

Another extension of the classes $\mathcal{P}$ and $\mathcal{NP}$ uses oracles as a source of computational power rather than as a source of information. For a given language A we shall consider *oracle Turing machines* that, during the computation, may ask queries of the form: "is $x \in A$?". Here, the oracle A is considered as part of the model of computation, rather than as part of the input. This notion of an oracle can help, for example, in understanding the relative difficulty of some problems.

For a language A, we denote by $\mathcal{P}^A$ and $\mathcal{NP}^A$ the *relativized* analogs of $\mathcal{P}$ and $\mathcal{NP}$ that are defined by Turing machines that use an oracle that decides membership in A. In general, the relativized analog of a complexity class c is denoted by c^A. The main result concerning oracle complexity classes, due to Baker, Gill and Solovay (1975), is that the answer to the relativized version of the $\mathcal{P} = \mathcal{NP}$ problem depends on the oracle. The intuition for the first alternative of the following theorem is given by the case when the oracle is an input. This is made rigorous by diagonalizing over all oracle Turing machines to construct an oracle A such that the language $L(A) = \{1^n\colon \exists x \in A \text{ such that } |x| = n\}$ is not in $\mathcal{P}^A$.

Theorem 2.5. *There exist languages A and B such that* $\mathcal{P}^A \neq \mathcal{NP}^A \neq \text{co-}\mathcal{NP}^A$ *and* $\mathcal{P}^B = \mathcal{NP}^B = \text{co-}\mathcal{NP}^B$.

Several of the theorems and proof techniques discussed in this survey easily extend to relativized complexity classes (i.e., they *relativize*). As a corollary to the above theorem, we can assert that techniques that relativize cannot settle the $\mathcal{P}$ versus $\mathcal{NP}$ problem.

One might wonder which one of the two alternatives provided by the theorem of Baker et al. is more typical. We define a random language A to be one that contains each word x independently with probability $\frac{1}{2}$. We say that a statement *holds for a random oracle* if the probability that the statement holds for a random language A in place of the oracle is 1. The Kolmogorov 0–1 law of probability theory states that if an event $\mathcal{A}$ is determined by the independent events, $\mathcal{B}_1$, $\mathcal{B}_2, \ldots$ and $\mathcal{A}$ is independent of any event that is a finite combination of the events $\mathcal{B}_i$, then the probability of $\mathcal{A}$ is either 0 or 1. Applying this to the event $\mathcal{A}$ that $\mathcal{P}^A = \mathcal{NP}^A$ and the events $\mathcal{B}_i$ that the ith word is in the language A, we see that the probability of $\mathcal{P}^A = \mathcal{NP}^A$ for a random oracle A is either 0 or 1. Bennett and Gill have provided the answers to these questions.

Theorem 2.6. *$\mathcal{P}^A \neq \mathcal{NP}^A$ and $\mathcal{NP}^A \neq$ co-$\mathcal{NP}^A$ for a random oracle A.*

2.4. *Evidence of intractability: $\mathcal{PSPACE}$-completeness*

In this subsection, we turn to the question of the space complexity of problems. When we discuss space complexity, we may assume that the Turing machine has only one work tape (and a separate input tape), since by expanding the tape alphabet, any number of tapes can be simulated by one tape without using more space. We shall also assume that the Turing machine halts in a unique configuration when accepting the input, e.g., it erases its work tape, moves both heads to the beginning of the tapes, and enters a special accepting state.

We remarked when introducing $\mathcal{PSPACE}$ that it was not an oversight that $\mathcal{NPSPACE}$ was not defined, since $\mathcal{PSPACE} = \mathcal{NPSPACE}$. This result is a special case of the following theorem of Savitch (1970).

Theorem 2.7. *If L is accepted by a nondeterministic Turing machine using space $S(n) \geqslant \log n$, then it is accepted by a deterministic Turing machine using space $S(n)^2$.*

Proof. The proof is based on the idea of modeling the computation by a directed reachability problem and using a natural divide-and-conquer strategy. In any given computation, a nondeterministic Turing machine M can be completely described by specifying the input head position, the contents of the work tape, the work-tape head position and the current state. Consider the directed graph G whose nodes correspond to these configurations and whose arcs correspond to possible transitions of M. The Turing machine accepts the input if and only if there is a path in G from the starting configuration to the (unique) accepting configuration.

Since M uses $S(n)$ space, there are at most $d^{S(n)}$ configurations for some constant d and hence the length of a simple path in G is at most $d^{S(n)}$. To solve the reachability problem in G, we build a procedure that recursively calls itself to check for any nodes i and k of G, and a bound l, whether there is a path from i to k of length at most 2^l. There is such a path if there exists a midpoint of the path j such that j can be reached from i, and k can be reached from j by paths of length at most 2^{l-1}. The existential quantifier can be implemented by merely trying all possible nodes j in some specified order. The basis of this recursion is the case $l = 0$, where we merely need to know if $i = k$, or if i and k are connected by an arc of G, and this can be easily checked in $O(S(n))$ space. In implementing this procedure, we need to keep track of the current midpoint at each level of the recursion, and so we need $lS(n)$ space to do this. To simulate M by a deterministic Turing machine, we run the procedure for the graph G with $l = S(n) \log d$ and the nodes corresponding to the starting and the accepting configurations. □

A problem L_1 is called *$\mathcal{PSPACE}$-complete* if $L_1 \in \mathcal{PSPACE}$ and for all $L \in \mathcal{PSPACE}$, $L \propto_p L_1$. The simplest $\mathcal{PSPACE}$-complete problem is the problem of determining if two nodes are connected in a directed graph G (of order 2^n) that is given by a circuit with $2n$ inputs where the first n specify in binary a node i and the second n specify a node j, and the output of the circuit is 1 if and only if (i,j) is an arc of G. This problem can easily be solved by a non-deterministic Turing machine using polynomial space hence it is in $\mathcal{PSPACE}$. To prove that it is complete, consider a language $L \in \mathcal{PSPACE}$. L can be reduced to this *circuit-based directed reachability problem* by introducing the graph G used in the proof of Savitch's theorem. It is easy to construct in polynomial time a (polynomial-size) circuit that decides if one configuration follows another by one transition of M.

The idea of "guessing a midpoint" is also the main idea used to derive another $\mathcal{PSPACE}$-complete problem. The problem of the validity of quantified Boolean formulae is as follows: given a formula in first-order logic in prenix form, $\exists x_1 \forall x_2 \cdots Q_k x_k \phi(x_1, \ldots, x_k) = \text{true}$, decide if it is valid. This problem is clearly in $\mathcal{PSPACE}$. The proof of its completeness is a mixture of Theorem 2.7 and Theorem 2.3. We use the alternation of existential and universal quantifiers to capture the notion of the existence of a midpoint such that both (for all) the first and second halves of the computation are legitimate. The basis of the recursion is now solved by building a Boolean formula to express that either two configurations are the same or one is the result of a single transition from the other.

An instance of the previous problem can be viewed as a game between an existential player and a universal player; the existential player gets to choose values for x_1, then the universal player chooses for x_2, and so on. The decision question amounts to whether the first player has a strategy such that ϕ must evaluate to true. There are many other $\mathcal{PSPACE}$-complete problems known, and most of these have a game-like flavor. An example of a more natural $\mathcal{PSPACE}$-complete game is the *directed Shannon switching game*: given a graph G with two nodes s and t, two players alternately color the edges of G, where the first player, coloring with red, tries to construct a red path from s to t, whereas the second player, coloring with blue, tries to construct a blue (s,t)-cut; does the first player have a winning strategy for G? Note that this result is in stark contrast to the undirected case, which can be solved in polynomial time (see chapter 11).

The role of games in $\mathcal{PSPACE}$-completeness suggests a new type of Turing machine, called an *alternating Turing machine*, which was originally proposed by Chandra, Kozen and Stockmeyer (1981). Consider a computation to be a sequence of moves made by two players, an existential player and a universal player. The current state indicates whose move it is, and in each configuration the specified player has several moves from which to choose. Each computation path either accepts or rejects the input. The input is accepted if the existential player has a winning strategy, that is, if there is a choice of moves for the existential player, so that for any choice of moves by the universal player, the input is accepted. For simplicity, assume again that each computation path has the same number of moves. As before, the time to accept an input x is this number of moves, and the space needed to accept x is the maximum space used on any computation path. Observe that a nondeterministic machine is an alternating machine where only the existential player moves.

The role of $\mathcal{PSPACE}$ in the definition of these machines suggests that alternating polynomial time is closely related to $\mathcal{PSPACE}$ and indeed this is just a special case of a general phenomenon. Let ATIME($T(n)$) and ASPACE($S(n)$) denote the classes accepted by an alternating Turing machine with $\mathrm{O}(T(n))$ time and $\mathrm{O}(S(n))$ space, respectively. Chandra, Kozen and Stockmeyer proved two fundamental results characterizing the relationship of alternating classes to deterministic and nondeterministic ones. Note that the first result, in essence, implies Savitch's theorem, and in fact, the proof of the last inclusion of Theorem 2.8 is very similar to the proof of Savitch's theorem.

Theorem 2.8. *If $T(n) \geqslant n$, then*

$$\mathrm{ATIME}(T(n)) \subseteq \mathrm{DSPACE}(T(n)) \subseteq \mathrm{NSPACE}(T(n)) \subseteq \mathrm{ATIME}(T(n)^2).$$

Theorem 2.9. *If $S(n) \geqslant \log n$, then* $\mathrm{ASPACE}(S(n)) = \bigcup_{c>0} \mathrm{DTIME}(c^{S(n)})$.

Among the consequences of these results, we see that $\mathcal{AP} = \mathcal{PSPACE}$. One can view an alternating Turing machine as an extremely idealized parallel computer, since it can branch off an unbounded number of parallel processes that can be used in determining the final outcome. Therefore, one can consider these results as a proven instance of the *parallel computation thesis*: parallel time equals sequential space (up to a polynomial function).

2.5. *The polynomial-time hierarchy*

The definition of $\mathcal{PSPACE}$ using alternating Turing machines suggests a hierarchy of complexity classes between $\mathcal{NP}$ and $\mathcal{PSPACE}$, called the *polynomial-time hierarchy*. The kth level of the polynomial-time hierarchy can be defined in terms of polynomial-time alternating Turing machines where the number of "alternations" between existential and universal quantifiers is less than k. Equivalently, we can define

$$\Sigma_k^{\mathcal{P}} = \{L \mid \exists L' \in \mathcal{P} \text{such that } x \in L \text{ if and only if}$$
$$\exists_p y_1 \forall_p y_2 \cdots Q_k y_k (x, y_1, y_2, \ldots, y_k) \in L'\},$$

where Q_k is $\exists_p$ if k is odd, and $\forall_p$ if k is even. Note that $\Sigma_0^{\mathcal{P}} = \mathcal{P}$ and $\Sigma_1^{\mathcal{P}} = \mathcal{NP}$. We also define $\Pi_k^{\mathcal{P}}$ to be co-$\Sigma_k^{\mathcal{P}}$. Clearly, $\Sigma_k^{\mathcal{P}} \cup \Pi_k^{\mathcal{P}} \subseteq \Sigma_{k+1}^{\mathcal{P}}$. The following *generalized coloring problem* gives a natural example of a problem in $\Sigma_2^{\mathcal{P}}$. Given input graphs G and H, can we color the nodes of G with two colors so that the graph induced by each color does not contain H as a subgraph?

Alternatively, it is possible to define the polynomial-time hierarchy in terms of oracles, as was done in the original formulation by Meyer and Stockmeyer (1972). For complexity classes $\mathscr{C}$ and $\mathscr{D}$ let $\mathscr{C}^{\mathscr{D}}$ denote the union of $\mathscr{C}^A$ over all $A \in \mathscr{D}$. Consider again the generalized coloring problem; it is not hard to see that there is a nondeterministic polynomial-time Turing machine to solve it, given an oracle for the following problem A: given G and H, decide if H is a subgraph of G. Since $A \in \mathscr{NP}$, we see that this coloring problem is in $\mathscr{NP}^{\mathscr{NP}}$. In fact, $\Sigma_2^{\mathscr{P}} = \mathscr{NP}^{\mathscr{NP}}$ and in general, $\Sigma_{k+1}^{\mathscr{P}} = \mathscr{NP}^{\Sigma_k^{\mathscr{P}}}$. Unfortunately, for each new complexity class, there is yet another host of unsettled questions.

Open Problems. For each $k \geqslant 1$, is $\Sigma_k^{\mathscr{P}} = \Sigma_{k-1}^{\mathscr{P}}$? For each $k \geqslant 1$, is $\Sigma_k^{\mathscr{P}} = \Pi_k^{\mathscr{P}}$?

Contained in these, for $k = 1$, are the $\mathscr{P} = \mathscr{NP}$ and $\mathscr{NP} = \text{co-}\mathscr{NP}$ questions, and as was true for those questions, one might hope to find complete problems on which to focus attention in resolving these open problems. We define these notions of completeness with respect to polynomial-time reducibility, so that L is complete for $\mathscr{C}$ if and only if $L \in \mathscr{C}$ and all problems in $\mathscr{C}$ reduce ($\propto_{\mathrm{p}}$) to L. As might be expected, analogs of the satisfiability problem which allow a particular number of alternations in the formulae can be used to provide a complete problem for each level of the hierarchy. On the other hand, it is more satisfying to have more natural complete problems, and Rutenberg showed that the generalized coloring problem is, in fact, complete for $\Sigma_2^{\mathscr{P}}$. Another problem of identical complexity is a similarly flavored node-deletion problem: given graphs G and H and an integer k, decide if there is a subset of k nodes that can be deleted from G, so that the remaining graph no longer has H as a subgraph.

One piece of good news concerning this infinite supply of open problems, is that their answers may be related. There is a principle of upward inheritability that says that if $\Sigma_l^{\mathscr{P}} = \Pi_l^{\mathscr{P}}$ for some level l, then equality holds for all levels $k \geqslant l$. In fact, $\Sigma_l^{\mathscr{P}} = \Pi_l^{\mathscr{P}}$ implies that the entire hierarchy collapses to that level; i.e., $\Sigma_k^{\mathscr{P}} = \Sigma_l^{\mathscr{P}}$ for all $k \geqslant l$. Note that $\mathscr{P} = \mathscr{NP}$ if and only if $\mathscr{P} = \mathcal{PH}$, where $\mathcal{PH} = \bigcup_{k \geqslant 0} \Sigma_k^{\mathscr{P}}$.

As we shall see, the polynomial-time hierarchy has helped to provide insight into the structure of several complexity classes. Perhaps the first result along these lines is due to Sipser, who used a beautiful "hashing function" technique to show that $\mathcal{BPP}$ is in the polynomial-time hierarchy, and in fact, can be placed within $\Sigma_2^{\mathscr{P}} \cap \Pi_2^{\mathscr{P}}$.

Furst, Saxe and Sipser (1981) discovered an interesting connection between constant-depth circuit lower bounds and separating relativized complexity classes. In particular, they showed that there are important consequences of an exponential lower bound on the size of a constant-depth circuit for the *parity function*, the sum modulo 2 of the bits of the input. Based on earlier results of Furst, Saxe and Sipser and Ajtai, Yao (1985) proved a sufficiently strong lower bound to yield the following theorem.

Theorem 2.10. *There exists an oracle A that separates the polynomial-time hierarchy from $\mathcal{PSPACE}$; that is, $\bigcup_{k \geqslant 1} \Sigma_k^{\mathscr{P},A} \neq \mathcal{PSPACE}^A$.*

The idea of the proof is as follows. First one shows that it is sufficient to consider alternating Turing machines in which every branch of the computation has a single oracle question at the end of the branch. The computation tree of such an alternating Turing machine with k levels of alternation corresponds to a depth-k circuit where the oracle answers are the inputs. For any oracle A, we define the language $L(A) = \{1^n \mid A$ contains an odd number of strings of length $n\}$. Now, $L(A)$ is in $\mathcal{PSPACE}^A$ for any oracle A. Using diagonalization and the result that for any constant c, a constant-depth circuit that computes the parity function has $\Omega(n^{\log^c n})$ gates, one can construct an oracle A such that $L(A)$ is not in $\bigcup_{k \geqslant 1} \Sigma_k^{\mathcal{P},A}$.

Another related problem is to separate the finite levels of the polynomial-time hierarchy using oracles. Baker and Selman proved the existence of an oracle A such that $\Sigma_2^{\mathcal{P},A} \neq \Pi_2^{\mathcal{P},A}$ (and consequently, $\Sigma_2^{\mathcal{P},A} \neq \Sigma_3^{\mathcal{P},A}$). Sipser showed that an oracle that separates finite levels of the polynomial-time hierarchy can be obtained via a connection to lower bounds on the size of constant-depth circuits, similar to the one used in Theorem 2.10. The following theorem is based on this as well as the requisite lower bounds for a certain function F_k, which is in some sense the generic function that is computable in depth k; these lower bounds were obtained through a series of results by Sipser, Yao and Hastad.

Theorem 2.11. *For every k, there exists an oracle A_k such that $\Sigma_k^{\mathcal{P},A_k} \neq \Sigma_{k+1}^{\mathcal{P},A_k}$.*

Techniques for proving the circuit-complexity lower bounds used in Theorems 2.10 and (2.11) are discussed in chapters 33 and 40. Based on stronger circuit-complexity results, Babai and Cai separated $\mathcal{PSPACE}$ from the finite levels of the hierarchy by random oracles. The question as to whether random oracles separate the finite levels of the polynomial-time hierarchy remains open. Another open question concerning random oracles is whether $\mathcal{P}^A = \mathcal{NP}^A \cap \text{co-}\mathcal{NP}^A$ for a random oracle A.

2.6. *Evidence of intractability: #$\mathcal{P}$-completeness*

Consider the *counting problem* of computing the number of perfect matchings in a bipartite graph. If this is cast as a decision problem, it does not appear to be in $\mathcal{NP}$, since it seems that the number of solutions can only be certified by writing down possibly exponentially many matchings. However, consider modifying the definition of $\mathcal{NP}$ to focus on the *number* of good certificates, rather than the existence of one; let #$\mathcal{P}$ be the class of problems for which there exists a language $L' \in \mathcal{P}$ and a polynomial $p(n)$ such that for any input x, the only acceptable output is $z = |\{y\colon |y| = p(|x|)$ and $(x,y) \in L'\}|$. Clearly, the problem of counting perfect matchings in a bipartite graph is in #$\mathcal{P}$. Any problem in #$\mathcal{P}$ can be computed using polynomial space, since all possible certificates y may be tried in succession. A recent result of Toda gives evidence of the intractability of this class by showing that $\mathcal{PH} \subseteq \mathcal{P}^{\#\mathcal{P}}$.

We can also define a notion of a complete problem for #$\mathcal{P}$. To do this, we use a reduction analogous to that used by Cook. Let P_1 and P_2 be counting problems, where the outputs for input x are denoted by $P_1(x)$ and $P_2(x)$, respectively. The problem P_1 reduces to P_2 if there exists a polynomial-time Turing machine that can compute P_1 if it has access to an oracle that, given x, can compute $P_2(x)$ in one step. We define a counting problem to be *#$\mathscr{P}$-complete* if it is in #$\mathscr{P}$, and all problems in #$\mathscr{P}$ reduce to it. A stronger notion of reducibility, analogous to the notion used by Karp, is called *parsimonious*: an instance x of P_1 is mapped in polynomial time to $f(x)$ for P_2, such that $P_1(x) = P_2(f(x))$. For either notion of reducibility, we see that any polynomial-time algorithm for P_2 yields a polynomial-time algorithm for P_1. It is not hard to see that the proof of Cook's theorem shows that the problem of computing the number of satisfying assignments of a Boolean formula in conjunctive normal form is #$\mathscr{P}$-complete. Furthermore, by being only slightly more careful, it is easy to give parsimonious modifications of the reductions to the clique problem, the hamiltonian circuit problem, and seemingly any $\mathscr{N}\mathscr{P}$-complete problem, and so the counting versions of $\mathscr{N}\mathscr{P}$-complete problems can be shown to be #$\mathscr{P}$-complete.

Surprisingly, not all #$\mathscr{P}$-complete counting problems need be associated with an $\mathscr{N}\mathscr{P}$-complete problem. Computing the number of perfect matchings in a bipartite graph (or equivalently, computing the *permanent* of a (0,1)-matrix) is a counting version of a problem in $\mathscr{P}$, the perfect matching problem, and yet Valiant (1979) has shown that this problem is #$\mathscr{P}$-complete. This made it possible to prove that a variety of counting problems are #$\mathscr{P}$-complete although they are based on polynomially solvable decision problems.

There are many problems that are essentially #$\mathscr{P}$-complete, although they do not have the appearance of a counting problem. An important example is the problem of computing the volume of a convex body. In the special case when the body is a polytope given by a system of linear inequalities, this was shown to be #$\mathscr{P}$-hard by Khachiyan and by Dyer and Frieze. For some problems, computing a good estimate suffices, and this might be much easier. If the convex body is given by an oracle that answers whether a given point is feasible, and if not produces a separating hyperplane (see chapter 28), then it is most natural to only estimate the volume. However, Bárány and Füredi, extending work of Elekes, proved that if U and L are upper and lower bounds on the volume of a d-dimensional convex body that are computed by an algorithm that makes only a polynomial number of calls to the oracle, then $U/L = \Omega((d/\log d)^d)$. Surprisingly, randomization can overcome this intractability. Dyer, Frieze and Kannan showed that there is a randomized algorithm that, given any $\varepsilon, \delta > 0$, computes upper and lower bounds U and L such that $U/L \leqslant 1 + \varepsilon$, uses a number of calls to the oracle that is bounded by polynomial in d, $1/\varepsilon$, and $\log(1/\delta)$, and is correct with probability at least $1 - \delta$.

Another problem that has been shown to be intimately connected with the estimation of the value of counting problems is that of uniformly generating combinatorial structures. As an example, suppose that a randomized algorithm requires a perfect matching in a bipartite graph G to be chosen uniformly from the set of all perfect matchings in G; how can this be done? Jerrum, Valiant and Vazirani have shown that a relaxed version of this problem, choosing perfect matchings that are selected with probability within an arbitrarily small constant factor of the uniform value, is equivalent to the problem of estimating the number of perfect matchings (using a randomized algorithm). In fact, their result carries through to most counting/generation problems related to problems in $\mathcal{NP}$, since it requires only a natural self-reducibility property that says that an instance can be solved by handling some number of smaller instances of the same problem.

Stockmeyer has provided insight into estimating the value of $\#\mathcal{P}$ problems, by trying to place these problems within the polynomial-time hierarchy. Using Sipser's hashing function technique, Stockmeyer showed that for any problem in $\#\mathcal{P}$ and any fixed value of $d > 0$, there exists a polynomial-time Turing machine with a $\Sigma_2^{\mathcal{P}}$-complete oracle that computes a value that is within a factor of $(1 + n^{-d})$ of the correct answer.

2.7. *Proof of intractability*

It is, of course, far preferable to *prove* that a problem is intractable, rather than merely giving evidence that supports this belief. Perhaps the first natural question is, are there *any* decidable languages that require more time than $T(n)$? The diagonalization techniques used to show that there are undecidable languages can be used to show that for any Turing-computable $T(n)$, there must exist such a language; consider the language L of all i such that if i is run on the Turing machine M_i that i encodes, either it is rejected, or it runs for more than $T(n)$ steps. It is easy to see that L is decidable, and yet no Turing machine that always halts within $T(n)$ steps can accept L. Using our stronger assumption about $T(n)$ (full time-constructibility) we are able to define such a language that is not only decidable, but can be recognized within only slightly more time than $T(n)$. The additional logarithmic factor needed in Theorem 2.12, which combines a result of Hartmanis and Stearns (1965) with one of Hennie and Stearns, is due to the fact that the fastest known simulation of a (multitape) Turing machine by a machine with some constant number of tapes slows down the machine by a logarithmic factor.

Theorem 2.12. *If*

$$\liminf_{n\to\infty} \frac{T_1(n)\log T_1(n)}{T_2(n)} = 0,$$

then there exists $L \in \mathrm{DTIME}(T_2) \setminus \mathrm{DTIME}(T_1)$.

Since any multitape machine can be simulated by a 1-tape machine without using any additional space, the corresponding space-hierarchy theorem of Hartmanis, Lewis and Stearns does not require the logarithmic factor. Seiferas, Fischer and Meyer have proved an analogous, but much more difficult, hierarchy theorem for nondeterministic time.

Theorem 2.13. *If*

$$\liminf_{n\to\infty} \frac{T_1(n+1)}{T_2(n)} = 0,$$

then there exists $L \in \mathrm{NTIME}(T_2) \setminus \mathrm{NTIME}(T_1)$.

Although the proofs of these theorems are non-constructive, Meyer and Stockmeyer developed the following strategy that makes use of completeness results in order to prove lower bounds on particular problems. Intuitively, a completeness result says that a problem is a "hardest" problem for some complexity class, and Theorems 2.12 and 2.13 can be used to show that certain complexity classes have provably hard problems. Consequently, these two pieces imply that the complete problem is provably hard.

As an example, consider the *circuit-based large clique problem*, L_{cc}, which is the problem analogous to the circuit-based directed reachability problem, that tests whether the (compactly represented) input graph on $N = 2^n$ nodes has a clique of size $N/2$. This problem is complete for the class $\mathcal{NEXPTIME} = \bigcup_k \mathrm{NTIME}(2^{n^k})$ with respect to polynomial-time reducibility. This can be seen by introducing the circuit-based version of satisfiability; a formula is represented by a polynomial-size circuit that outputs 1 for input (i, j) when literal l_i is in the jth clause, where i and j are given in binary. The generic reduction of Cook's theorem translates exactly to show that circuit-based satisfiability is $\mathcal{NEXPTIME}$-complete, and the completeness of L_{cc} follows by using essentially the same reduction used to show the $\mathcal{NP}$-completeness of the ordinary clique problem. Now consider a language $L \in \mathrm{NTIME}(2^{n^2}) - \mathrm{NTIME}(2^n)$ specified by Theorem 2.13. Since $L \propto_{\mathrm{p}} L_{\mathrm{cc}}$, then $L_{\mathrm{cc}} \in \mathrm{NTIME}(2^{n^c})$ implies that $L \in \mathrm{NTIME}(p(n) + 2^{p(n)^c})$, where $p(n) = n^k$ is the time bound for the reduction. By choosing $c = 1/k$, we obtain the following theorem.

Theorem 2.14. *There exists a constant* $c > 0$ *such that* $L_{\mathrm{cc}} \notin \mathrm{NTIME}(2^{n^c})$.

One interpretation of this result is that there exist graphs specified by circuits such that proofs that these graphs have a large clique require an exponential number of steps (in terms of the size of the circuit). Observe that we would have been able to prove a stronger result if we had had a better bound on the length of the string produced by the reduction. Lower bounds proved using this strategy are often based on completeness with respect to reducibilities that are further restricted to produce an output of length bounded by a linear function of the input length.

This strategy has been applied primarily to problems from logic and formal language theory. A result of Fischer and Rabin (1974) that contrasts nicely with Theorem 1.3 concerns L_{pa}, the language of all provable sentences in the theory of arithmetic for natural numbers without multiplication, which was shown to be decidable by Presburger.

Theorem 2.15. *There is a constant* $c > 0$ *such that* $L_{\mathrm{pa}} \notin \mathrm{NTIME}(2^{2^{cn}})$.

A representative sample of problems treated in this fashion is surveyed by Stockmeyer (1987).

There are two other important results that separate complexity classes. Hopcroft, Paul and Valiant (1977) showed that $\mathrm{DTIME}(T(n)) \subseteq \mathrm{DSPACE}(T(n)/\log T(n))$, and thus time is not the same complexity measure as space. Paul, Pippenger, Szemerédi and Trotter (1983) showed that $\mathrm{DTIME}(n) \neq \mathrm{NTIME}(n)$.

2.8. Extensions of $\mathcal{NP}$: Short proofs via randomization

In the same way that randomized algorithms give rise to an extended notion of efficiently computable, randomized proofs give rise to an extension of $\mathcal{NP}$, the class of languages for which membership is efficiently provable. Randomized proofs give overwhelming statistical evidence. In creating this branch of complexity theory, Babai (1985) and Goldwasser, Micali and Rackoff (1989) have given definitions to capture related notions of proof based on statistical evidence. The interested reader is directed to the surveys by Goldwasser (1989) and Johnson (1988).

Suppose that King Arthur has two graphs G_1 and G_2, and his magician Merlin wants to convince him that the two graphs are not isomorphic. The graph isomorphism problem is not known to be in co-$\mathcal{NP}$, but Goldreich, Micali and Wigderson have given the following way for Merlin to convince Arthur that the two graphs are not isomorphic. Merlin asks Arthur to choose one of the two graphs, randomly relabel the nodes, and show him this random isomorphic copy of the chosen graph; Merlin will tell which graph Arthur chose. If the two graphs are isomorphic, then Merlin has only a fifty percent chance of choosing the right graph, assuming that he cannot read Arthur's mind. If Merlin can successfully repeat this test several times, then Arthur can fairly safely conclude that Merlin can distinguish isomorphic copies of the two graphs; in particular, the two graphs are not isomorphic.

The above randomized proof is a prime example of the interactive proof defined by Goldwasser, Micali and Rackoff. An *interactive protocol* consists of two Turing machines: the Prover (Merlin) which has unrestricted power and the Verifier (Arthur) which is restricted to be a randomized polynomial-time Turing machine. The two machines operate in turns, and communicate only between turns via a special tape. The Prover is trying to convince the Verifier that a common input x is in a language L. A language L has an *interactive proof system* if there exists an interactive protocol such that if $x \in L$, then the Verifier accepts with probability at least $\frac{2}{3}$; if $x \notin L$, then for any Turing machine used in place of the Prover, the Verifier rejects with probability at least $\frac{2}{3}$. Let $\mathcal{IP}$ denote the class of languages that have an interactive proof system with a polynomial number of turns. Note that, once again, the choice of probability $\frac{2}{3}$ is arbitrary.

Babai has proposed a similar, but seemingly weaker, version of a randomized proof system, an *Arthur–Merlin game*, in trying to capture a complexity class just above $\mathcal{NP}$. It can be defined in the same way as an interactive proof system with the assumption that each machine may access the other's work and randomizing tapes. Note the importance of privacy in the above interactive protocol. Goldwasser and Sipser proved, however, that interactive proofs and Arthur–Merlin games define the same complexity class.

One can define randomized proof hierarchies in a way analogous to the polynomial-time hierarchy. We consider the class of languages accepted by an interactive proof system (or an Arthur–Merlin game) with less than k alternations of turns. Babai introduced $\mathcal{AM}$ to denote the class of languages that have an Arthur–Merlin game where a single move of Arthur is followed by a single move of Merlin. It would be natural to conjecture that these hierarchies are strict. However, Babai proved that if a problem has an Arthur–Merlin game with a finite number of turns than it is in $\mathcal{AM}$. Since the equivalence between interactive proofs and Arthur–Merlin games increases the number of turns only by a constant, the same is true for interactive proofs with a finite number of turns.

On the other hand, it appears that $\mathcal{IP}$, which allows a polynomially bounded number of alternations, is a significantly richer class than $\mathcal{AM}$. Lund, Fortnow, Karloff and Nisan showed that $\mathcal{PH} \subseteq \mathcal{IP}$, and Shamir extended their techniques to prove the following theorem.

Theorem 2.16. $\mathcal{IP} = \mathcal{PSPACE}$.

One way to view this result is that it is possible to convince someone of a theorem in polynomial time, if it can be proven using a polynomial-sized blackboard. An interesting aspect of these results is that they do not relativize, since, for example, Fortnow and Sipser have constructed an oracle A for which co-$\mathcal{NP}^A$ is not contained in $\mathcal{IP}^A$. There is evidence that $\mathcal{AM}$ is a more restrictive class. Just as $\mathcal{BPP} \subseteq \mathcal{P}$/poly, one can show that $\mathcal{AM}$ is contained in $\mathcal{NP}$/poly, a non-uniform extension of $\mathcal{NP}$. Babai has shown that $\mathcal{AM} \subseteq \Pi_2^{\mathcal{P}}$ by extending the proof that $\mathcal{BPP} \subseteq \Sigma_2^{\mathcal{P}} \cap \Pi_2^{\mathcal{P}}$. It appears that $\mathcal{AM}$ does not contain co-$\mathcal{NP}$, but to prove this would imply that $\mathcal{NP} \neq$ co-$\mathcal{NP}$. Nonetheless, the following theorem, due to Boppana, Hastad and Zachos, provides some evidence.

Theorem 2.17. *If* co-$\mathcal{NP} \subseteq \mathcal{AM}$ *then* $\Sigma_2^{\mathcal{P}} = \Pi_2^{\mathcal{P}} = \mathcal{AM}$.

We have seen that the graph non-isomorphism problem is in $\mathcal{AM}$. Therefore, Theorem 2.17 implies that if the graph isomorphism problem is $\mathcal{NP}$-complete, then $\Sigma_2^{\mathcal{P}} = \Pi_2^{\mathcal{P}} = \mathcal{AM}$.

Goldwasser, Micali and Rackoff introduced the interactive proof system in order to characterize the minimum amount of "knowledge" that needs to be transferred in a proof. Interactive proofs make it possible to "prove", for example, that two graphs are isomorphic, without giving any further clue about the isomorphism between them. These aspects of interactive proofs will be discussed in chapter 40.

In the years since this survey was written, there have been quite a number of very significant developments beyond the results mentioned here. We have already mentioned the result that $\mathcal{IP} = \mathcal{PSPACE}$; *this result was proved just in time to be included in the final revised version sent to the publisher, but it has turned out that this was more a beginning than a conclusion. Ben-Or, Goldwasser, Kilian, and Wigderson introduced an analogous notion of* multiprover *interactive proofs in which the Verifier can be convinced by several independent Provers that cannot communicate with each other during the protocol. Fortnow, Rompel, and Sipser gave an alternate characterization of this class* $\mathcal{MIP}$, *which was used by Babai, Fortnow, and Lund to prove that* $\mathcal{MIP} = \mathcal{NEXPTIME}$. *Fiege, Goldwasser, Lovász, Safra, and Szegedy showed, using ideas from the proof that* $\mathcal{MIP} = \mathcal{NEXPTIME}$, *that there is* *a fundamental connection between randomized complexity classes and proving that certain optimization problems are hard to solve even approximately. Extensions of this result by Arora and Safra and subsequently Arora, Lund, Motwani, Sudan, and Szegedy, led to the ultimate result along these lines: $\mathcal{NP}$ is exactly the class of languages L for which there is the following type of* probabilistically checkable proof; *for any input x, the Verifier is given a certificate of polynomial length of which it may query* O(1) *bits based on* O(log |*x*|) *random coin flips; for each $x \in L$, there exists a certificate such that the Verifier always accepts; for each $x \notin L$, given any certificate the Verifier rejects with probability at least $\frac{1}{2}$. For a survey of these results, the reader is referred to* Johnson (1992).

3. Living with intractability

The knowledge that a problem is $\mathcal{NP}$-complete is little consolation for the algorithm designer who needs to solve it. Contrary to their theoretical equivalence, all $\mathcal{NP}$-complete problems are not equally hard from a practical perspective. In this section, we will examine two approaches to these intractable problems that, while not overcoming their inherent difficulty, make them appear more manageable. In this process, finer theoretical distinctions will appear between these problems and will help to explain the empirical evidence.

3.1. The complexity of approximate solutions

Most of the 21 $\mathcal{NP}$-complete problems in Karp's original paper are decision versions of optimization problems; this is also true for a great majority of the problems catalogued by Garey and Johnson (1979). Although the combinatorial nature of these problems makes it natural to focus on optimal solutions, for most practical settings in which these problems arise it is nearly as satisfying to obtain solutions that are guaranteed to be nearly optimal. In this subsection we will briefly survey the sorts of *performance guarantees* that can and cannot be obtained for particular combinatorial problems. For further discussion of the algorithmic techniques used in obtaining near-optimal solutions, the reader is referred to chapter 28. Throughout this section, OPT(I) will denote the optimal value of an instance I of a particular combinatorial optimization problem.

It is possible that deciding if OPT(I) $\leqslant k$ is $\mathcal{NP}$-complete, and yet a solution of value no more than OPT(I) + 1 can be computed in polynomial time. In fact, this is true for the *edge coloring problem*, where we are given an undirected simple graph G and an integer k, and we wish to color the edges with as few colors as possible so that no two edges incident to the same node receive the same color. Holyer showed that it is $\mathcal{NP}$-complete to decide if a cubic graph can be 3-edge-colored. On the other hand, Vizing proved that the minimum number of colors needed is at most one more than maximum degree of G, and his proof immediately yields a polynomial-time algorithm that uses at most OPT(I) + 1 colors. Since edge-coloring graphs with chromatic index 2 is trivial, this also yields an algorithm that always uses no more than $\frac{4}{3} \cdot \mathrm{OPT}(I)$; a polynomial-time algorithm with such an *absolute performance guarantee* is often called a $\frac{4}{3}$-approximation algorithm. On the other hand, it is easy to see that this is the best possible performance guarantee, unless $\mathcal{P} = \mathcal{NP}$. Suppose that there exists a ρ-approximation algorithm for edge coloring with $\rho < 4/3$. If this algorithm is run on a cubic graph that can be colored with three colors, then the algorithm must return a coloring that uses fewer than four colors; it returns an optimal coloring.

One type of strong approximation result is called a *fully polynomial approximation scheme*; this is a family of algorithms $\{A_\varepsilon\}$, where A_ε is a $(1+\varepsilon)$-approximation algorithm and the dependence of the running time on ε is bounded by a polynomial in $1/\varepsilon$. By solving rounded instances with only a limited number of significant digits, Ibarra and Kim gave such a scheme for the *knapsack problem*: given n pieces to be packed into a knapsack of size B, where piece j has size s_j and value v_j, pack a subset of pieces of total size $\leqslant B$ with maximum total value. Note that it is impossible to improve the dependence of the running time on ε to a polynomial in $\log(1/\varepsilon)$ since it is always possible to choose ε of polynomial length, such that $A_\varepsilon(I) - \mathrm{OPT}(I) < 1$, and thus by integrality, $A_\varepsilon(I) = \mathrm{OPT}(I)$. The same argument implies an important result of Garey and Johnson; if a problem is strongly $\mathcal{NP}$-complete, then there is no fully polynomial approximation scheme for it unless $\mathcal{P} = \mathcal{NP}$. If the running time of A_ε may depend arbitrarily on $1/\varepsilon$, it is sometimes possible to obtain such a *polynomial approximation scheme* for strongly $\mathcal{NP}$-complete problems. Hochbaum and Shmoys showed that this is the case for the following *machine-scheduling problem*: each of n jobs is to be scheduled on one of m machines, where job j takes time p_j on any machine, and the aim is to minimize the time by which all jobs are completed. In fact, the idea of studying the performance guarantees of heuristics for optimization problems was first proposed in the context of this problem by Graham (1966), who gave a 2-approximation algorithm.

It is sometimes too restrictive to focus on ρ-approximation algorithms. A good illustration of this is the *bin-packing problem*: given a collection of n pieces, where piece j has size s_j, how many bins of size B are needed to pack all of the pieces? Since it is easy to formulate the subset sum problem as a question of whether 2 bins suffice, we see that a ρ-approximation algorithm with $\rho < \frac{3}{2}$ would imply that $\mathcal{P} = \mathcal{NP}$. However, Johnson showed that a simple heuristic uses at most $\frac{11}{9} \cdot \mathrm{OPT}(I) + 4$ bins. This suggests that an *asymptotic performance guarantee* may also be interesting, where we consider the infimum of the absolute performance guarantee for instances with $\mathrm{OPT}(I) \geqslant k$ (as $k \to \infty$). It was a great surprise when Fernandez de la Vega and Lueker not only substantially improved this $\frac{11}{9}$ bound, but gave a polynomial approximation scheme with respect to asymptotic guarantees. Perhaps even more surprisingly, Karmarkar and Karp extended this to give such a scheme where the dependence of the running time on $1/\varepsilon$ was bounded by a polynomial.

If it is possible to scale up the data to generate an equivalent problem, such as using processing times $M \cdot p_j$ in the machine-scheduling problem, any distinction between the absolute and asymptotic performance guarantees disappears. For node coloring, the following "graph composition" accomplishes this scaling: take M copies of the graph, and make each pair of nodes in different copies adjacent. The $\mathcal{NP}$-completeness of 3-colorability again implies that an absolute performance guarantee better than $\frac{4}{3}$ is unlikely. In fact, Garey and Johnson (1976) use a more intricate composition technique to *increase* this ratio, and thereby prove that an asymptotic performance guarantee less than 2 would imply that $\mathcal{P} = \mathcal{NP}$.

For some problems, such as the *traveling salesman problem*, Gonzalez and Sahni observed that no constant performance guarantee is possible unless $\mathcal{P} = \mathcal{NP}$. In this example, where the aim is find a minimum-length hamiltonian circuit in a complete graph where edge e has length c_e, one can use a ρ-approximation algorithm to decide the hamiltonian circuit problem for a graph $G = (V, E)$; set $c_e = 1$ for $e \in E$ and $\rho|V| + 1$ otherwise. However, Christofides has given a $\frac{3}{2}$-approximation algorithm for instances that satisfy the triangle inequality.

For the great majority of problems, such as node coloring, there is neither a constant performance guarantee nor any evidence that such an algorithm does not exist. In the case of the maximum stable set problem, for which the best-known algorithm has performance guarantee little better than $O(n/\log n)$, there is some evidence that no polynomial approximation scheme exists. Garey and Johnson have given another graph composition technique to show that if performance guarantee ρ^2 is obtained, then this can be used to obtain a ρ-approximation algorithm. By repeatedly applying this technique, it is possible to convert any ρ-approximation algorithm, where ρ is a constant, into a polynomial approximation scheme. There are few other techniques that provide evidence for the intractability of computing near-optimal solutions. Recently, Papadimitriou and Yannakakis have proposed a complexity class, along with a notion of completeness, that attempts to characterize those problems that have a constant performance guarantee, but do not have a polynomial approximation scheme.

In the years since this survey was written, there have been dramatic advances in proving that certain problems are also hard to approximate. Feige, Goldwasser, Lovász, Safra, and Szegedy showed that unless $\mathcal{NP} \subset$ DTIME$(n^{\log \log n})$, there does exist a ρ-approximation algorithm for maximum stable set problem for any constant $\rho > 1$. Arora and Safra strengthened this to show that achieving such an approximation is $\mathcal{NP}$-hard, and this was strengthed even further by Arora, Lund, Motwani, Sudan, and Szegedy, who proved that there exists an $\varepsilon > 0$ such that there does not exist an n^{ε}-approximation algorithm unless $\mathcal{P} = \mathcal{NP}$. These techniques have yielded signficant results for other problems as well. Lund and Yannakakis proved a hardness-of-approximation result for the minimum node coloring problem identical to the one stated above for the maximum stable set problem. Arora, Lund, Motwani, Sudan, and Szegedy also showed that, unless $\mathcal{P} = \mathcal{NP}$, there does not exist a polynomial approximation scheme for any complete problem in the class $\mathcal{MAX\ SNP}$ proposed by Papadimitriou and Yannakakis. For example, a corollary of this result is that there does not exist a polynomial approximation scheme for *the traveling salesman problem with the triangle inequality, unless $\mathcal{P} = \mathcal{NP}$. For a survey of this research, the reader is referred to* Johnson (1992) *and* Shmoys (1995).

3.2. Probabilistic analysis of algorithms

One justified criticism of complexity theory is that it focuses on worst-case possibilities, and this may in fact be unrepresentative of practical reality. In this section, we will briefly indicate some of the results that are concerned with the probabilistic analysis of algorithms, where inputs are selected according to a specified probability distribution, and the average behavior is studied. Many related results are presented in chapter 6, and the reader is referred there, as well as to the surveys of Karp, Lenstra, McDiarmid and Rinnooy Kan (1985) and Coffman, Lueker and Rinnooy Kan (1988). We shall also sketch the main ideas of an analog of $\mathcal{NP}$-completeness, recently proposed by Levin (1986), to provide evidence that a problem is hard to solve in even a probabilistic sense.

For all of the problems mentioned in the subsection on the worst-case analysis of heuristics, it is possible to obtain much more optimistic results for the average-case analysis under the assumption that the input is drawn from a specified probability distribution. Unfortunately, these results rely heavily on the particular distribution selected, and an approach to the average-case analysis of algorithms that is insensitive to this, would be an important advance. As an example, for the traveling salesman problem with edge lengths that are independently and identically distributed (i.i.d.) uniformly over the interval [0,1], Karp has given a heuristic where the expected value of its relative error is $\mathrm{O}(n^{\ 1/2})$. On the other hand, the nodes may be selected i.i.d. uniformly in the unit square $[0, 1]^2$, and the length of edge is given by the Euclidean distance between its endpoints. In this case, Karp (1976) has given a different algorithm that has the stronger property that the relative error converges to 0 almost surely as $n \to \infty$. This result, which stimulated much work in the area of probabilistic analysis, is based in part on a result of Beardwood, Halton and Hammersley, which proves that, for instances selected as above, there exists a constant c such that $\mathrm{OPT}(I)/\sqrt{n} \to c$ almost surely.

Similar results are also known for such problems as node coloring and the hamiltonian circuit problem. This work grew out of the theory of random graphs of Erdős and Rényi, which is treated in chapter 6. A common way to choose a random graph is to include each possible edge independently with probability $\frac{1}{2}$. For the first problem, it is possible to prove that a simple greedy method is, in probability, a factor of 2 more than optimal. For the latter, it is possible to give an algorithm that always gives a correct answer and runs in expected polynomial time.

Although the probabilistic analysis of algorithms has focused mainly on $\mathcal{NP}$-complete problems, it has often served as a useful tool to show that the average-case running time of certain algorithms is much better than the worst-case running time. The most important example of this is the simplex method for linear programming, which is a practically efficient algorithm that was shown to have exponential worst-case running time by Klee and Minty. Borgwardt and Smale, independently showed that variants of the simplex method take polynomial expected time under certain probabilistic assumptions. For a thorough survey of results in this area, the reader is referred to Shamir (1987).

Not all problems in $\mathcal{NP}$ have been solved efficiently even with probabilistic assumptions, and only the simplest sorts of distributions have been analyzed. This raises the specter of intractability: are there distributions for which certain $\mathcal{NP}$-complete problems remain hard to solve? Levin (1986) has proposed a notion of completeness in a probabilistic setting. Once again, evidence for hardness is given by showing that if a particular problem in $\mathcal{NP}$ with a specified input distribution can be solved in expected polynomial time, then every such problem and distribution pair can also be solved so efficiently. For a more complete discussion, the reader is encouraged to read the column of Johnson (1984). One of the motivations for studying such truly intractable problems come from the area of cryptography, which attempts to use the intractability of a problem to the algorithm designer's advantage (see chapter 40).

A bit of care must be given to formulating the precise notion of polynomial expected time, so that it is insensitive to both the particular choice of machine and to the encoding. If $\mu(x)$ denotes the probability that a randomly selected instance of size n is x, and $T(x)$ is the running time on x, then one would typically define expected polynomial time to require that the sum of $\mu(x)T(x)$ over all instances of size n is $O(n^k)$ for some constant k. Instead, we consider μ to be the density function over the set of all instances I, and require that $\sum_{x \in I} \mu(x)T(x)^{1/k}/|x| < \infty$ for some constant k. Levin's notion of *random* $\mathcal{NP}$ requires that the distribution function $M(x) = \sum_{i=1}^{x} \mu(i)$ can be computed in $\mathcal{P}$, where each instance is viewed as a natural number. This notion does not seem to be too restrictive, and includes all of the probability distributions discussed here. It only remains to define the notion of reducibility. As usual, "yes" instances must be mapped by a polynomial-time function f to "yes" instances, and analogously for "no" instances, but one must consider the density functions as well. The pair (L_1, μ_1) reduces to (L_2, μ_2) if in addition we require that $\mu_2(x)$ is at least a polynomial fraction of the total probability of elements that are mapped to x by f.

Levin showed that all of random $\mathcal{NP}$ reduces to a certain *random tiling problem*. Instances consist of integers $k \leqslant N$, a set of tile types, each of which is a unit square labeled with letters in its four corners, and a side-by-side sequence of k such tiles where consecutive tiles have matching labels in both adjoining corners. We wish to decide if there is a way of extending the sequence to fill out an $N \times N$ square, where adjacent tiles have matching labels in their corresponding corners, and the top row starts with the specified sequence of tiles. The instances are selected by first randomly choosing a value for N, where N is set equal to n with probability proportional to n^{-2}, $n = 1, 2, \ldots$; k is chosen uniformly between 1 and N, the tile types are chosen uniformly, and then the tiles in the sequence are selected in order uniformly among all possible extensions of the current sequence. More recently, Venkatesan and Levin have shown that a generalization of the problem of edge-coloring digraphs (where for certain subgraphs, the number of edges given each color is specified) is also random $\mathcal{NP}$-complete.

4. Inside $\mathcal{P}$

In this section we shall focus on the complexity of problems in $\mathcal{P}$. After proving that a problem is in $\mathcal{P}$, the most important next step is undoubtedly to find an algorithm that is *truly* efficient, and not merely efficient in this theoretical sense. However, we will not address that issue, since it is best deferred to the individual chapters that discuss polynomial-time algorithms for particular problems. In this section, we shall address questions that relate to machine-independent complexity classes inside $\mathcal{P}$.

From a practical viewpoint, the most appealing complexity class inside $\mathcal{P}$ is the class of languages for which some polynomial-time algorithms can be speeded up significantly if several processors work simultaneously. We shall discuss parallel computation, and focus on the complexity class $\mathcal{NC}$, which serves as a theoretical model of efficient parallel computability, much as $\mathcal{P}$ serves as a theoretical model of efficient "sequential" computation.

We shall also consider the space complexity of problems in $\mathcal{P}$. Recall that the parallel computation thesis suggests that there is a close relationship between sequential space complexity and parallel time complexity. We will show another proven case of this thesis: every problem in $\mathcal{L}$, the class of problems solvable using logarithmic space, can be solved extremely efficiently in parallel. This is one source of interest in the complexity class $\mathcal{L}$. Another source is that this complexity class is the basis for the natural reduction that helps to distinguish between the problems in $\mathcal{P}$ in order to provide a notion of a hardest problem in $\mathcal{P}$.

4.1. *Logarithmic space*

The most general restriction on the space complexity of a language L that is known to imply that $L \in \mathcal{P}$ is logarithmic space. Observe that $\mathcal{L} \subseteq \mathcal{NL} \subseteq \mathcal{P}$, where the last inclusion follows, for example, from Theorem 2.9. The typical use of logarithmic space is to store a constant number of pointers, e.g., the names of a constant number of nodes in the input graph, and in some sense, this restriction attempts to characterize such algorithms. Although $\mathcal{L}$ contains many interesting examples, we see the role of logarithmic-space computation more as a natural means of reduction rather then providing interesting algorithms. Instead, we will focus on the nondeterministic and randomized analogs, for which there are languages that appear to be best characterized in terms of their space complexity.

To define the notion of a logarithmic-space reduction, we introduce a variant of logarithmic-space computation that can produce output of superlogarithmic size. We say that a function f can be *computed in logarithmic space* if there exists a Turing machine with a read-only input tape and a write-only output tape that, on input x, halts with $f(x)$ written on its output tape and uses at most logarithmic space on its work tapes. A problem L_1 *reduces in logarithmic space* to L_2 if there exists a function f computable in logarithmic space that maps instances of L_1 into instances of L_2 such that x is a "yes" instance of L_1 if and only if $f(x)$ is a "yes" instance of L_2. Let $L_1 \propto_{\log} L_2$ denote such a logarithmic-space reduction (or log-space reduction, for short). It is fairly easy to show that the $\propto_{\log}$ relation is transitive. A problem L_1 is *$\mathcal{NL}$-complete* if $L_1 \in \mathcal{NL}$ and for all $L \in \mathcal{NL}$, $L \propto_{\log} L_1$. The transitivity of $\propto_{\log}$ yields the following result.

Theorem 4.1. *For any $\mathcal{NL}$-complete problem L, $L \in \mathcal{L}$ if and only if $\mathcal{L} = \mathcal{NL}$.*

Savitch provided the most natural example of an $\mathcal{NL}$-complete language: the directed graph reachability problem. The problem is clearly in $\mathcal{NL}$ and can be shown to be $\mathcal{NL}$-complete along the same lines as the $\mathcal{PSPACE}$-completeness of the circuit-based directed reachability problem.

Theorem 4.2. *The directed-graph reachability problem is $\mathcal{NL}$-complete.*

In view of the above result, it was very surprising when Aleliunas, Karp, Lipton, Lovász and Rackoff (1979) showed that the *undirected-graph reachability problem*, the analog of the directed-graph reachability problem for undirected graphs, can be solved by a randomized Turing machine using logarithmic space. The algorithm attempts to find the required path by following a random walk in the graph, starting from the node s. It can be shown that a random walk is expected to use every edge with the same frequency, and if s and t are in the same connected component then the walk is expected to reach t in at most $O(nm)$ steps, where n and m denote the numbers of nodes and edges. We define the class $\mathcal{RL}$ to be the log-space analog of $\mathcal{RP}$. A language L is in $\mathcal{RL}$ if there exists a randomized Turing machine RM that works in logarithmic space, such that each input x that RM accepts along any computation path is in L and for every $x \in L$, the probability that RM accepts x is at least $\frac{1}{2}$.

Theorem 4.3. *Undirected-graph reachability is in $\mathcal{RL}$.*

Recently, another piece of evidence has been discovered that suggests that undirected graph reachability is easier than its directed analog. Nisan, Szemerédi, and Wigderson proved that the undirected graph reachability problem can be solved in DSPACE($\log^{1.5} n$).

One can think of the classes $\mathcal{L}$ and $\mathcal{NL}$ as lower-level analog of $\mathcal{P}$ and $\mathcal{NP}$. By studying the relationships of $\mathcal{L}$, $\mathcal{NL}$ and co-$\mathcal{NL}$, one hopes to better understand the relationship of deterministic and nondeterministic computation. It is this point of view that makes the following theorem of Immerman (1988) and Szelepcsényi (1987) one of the biggest surprises in recent developments in complexity theory.

Theorem 4.4. $\mathcal{NL}$ = co-$\mathcal{NL}$.

The proof uses a definition of nondeterministically computing a *function*. We say that a function $f(x)$ *can be computed in nondeterministic logarithmic space* if there is a nondeterministic log-space Turing machine that, on input x, outputs the value $f(x)$ on at least one branch of the computation and on every other branch either stops without an output or also outputs $f(x)$. If $f(x)$ is a Boolean function, then we say that the language L defined by $f(x) = 1$ is *decided in nondeterministic logarithmic space*, which is equivalent to L being in $\mathcal{NL} \cap$ co-$\mathcal{NL}$.

We will prove Theorem 4.4 by showing that the $\mathcal{NL}$-complete directed-graph reachability problem can be decided in nondeterministic logarithmic space. Given a directed graph $G = (V, A)$, a source node s and an integer k, let $f(G, s, k)$ denote the number of nodes reachable from the node s along paths of length at most k.

Lemma 4.5. *The directed-graph reachability problem is decidable in nondeterministic logarithmic space if and only if the function $f(G, s, k)$ can be computed in nondeterministic logarithmic space.*

Proof. To prove the *only if* direction, we use the fact that the directed graph reachability is $\mathcal{NL}$-complete. If it is decidable in logarithmic space, then so is the problem of recognizing if there is a path of length at most k. To compute $f(G, s, k)$, we use the assumed nondeterministic machine for each node v, to decide if v is reachable from s by a path of length at most k, and count the number of reachable nodes.

To prove the opposite direction, we use the following nondeterministic log-space computation. First compute $f(G, s, n)$. Then for each node v, (nondeterministically) try to guess a path from s to v. Count the number of nodes for which a path has been found. If a path has been found to t, we accept the input. If $f(G, s, n)$ nodes have been reached without finding a path to t, this proves that t is not reachable from s, so we reject. Finally, if the number of nodes that have been reached is less than $f(G, s, n)$, then this is an incorrect branch of the computation, and the computation stops without producing an output. □

To finish the proof of Theorem 4.4 we have to argue that the function $f(G, s, k)$ can be computed in nondeterministic logarithmic space. This is done by induction on k. Given $f(G, s, k)$ we can decide if there exists a path of length $k + 1$ from s to a particular node v by checking if there is a path of length at most k to any of the predecessors of v using a variant of the algorithm given in the *if* part of Lemma (4.5). Counting these nodes gives $f(G, s, k + 1)$.

4.2. *The hardest problems in $\mathcal{P}$*

One important application of log-space computation was introduced by Cook, who used log-space reducibility to introduce a notion of a hardest problem in $\mathcal{P}$. A problem L_1 is *$\mathcal{P}$-complete* if $L_1 \in \mathcal{P}$ and for all $L \in \mathcal{P}$, $L \propto_{\log} L_1$. The transitivity of the log-space reduction gives the following theorem.

Theorem 4.6. *For any $\mathcal{P}$-complete problem L, $L \in \mathcal{L}$ if and only if $\mathcal{L} = \mathcal{P}$.*

Later in this section we shall see that $\mathcal{P}$-completeness also provides evidence that a problem cannot be efficiently solved in parallel. This fact has greatly increased the interest in $\mathcal{P}$-completeness and a variety of problems have been shown to be $\mathcal{P}$-complete. Perhaps the most natural example is the circuit value problem, which was proved $\mathcal{P}$-complete by Ladner. Given a circuit with truth values assigned to its input gates, the *circuit value problem* is to compute the output of the circuit. This problem is clearly in $\mathcal{P}$. It can be proved $\mathcal{P}$-complete by building a circuit that simulates the computation of a Turing machine.

Theorem 4.7. *The circuit value problem is $\mathcal{P}$-complete.*

Dobkin, Lipton and Reiss proved that each problem in $\mathcal{P}$ log-space reduces to the linear programming problem, and the celebrated result of Khachiyan showed that it is in $\mathcal{P}$. Valiant gave a straightforward reduction from a restricted circuit value problem that uses linear constraints to trace the value computed by the circuit.

Goldschlager, Shaw and Staples showed that the *maximum flow problem*, an important special case of the linear programming problem, is also $\mathcal{P}$-complete. In this problem, we are given a directed graph $G = (V, A)$ with two specified nodes, the source s and the sink t, and a non-negative capacity $u(a)$ assigned to each arc $a \in A$. A feasible flow is a vector $f \in \mathbb{R}^A$ that satisfies the capacity constraints, i.e., $0 \leqslant f(a) \leqslant u(a)$ for each arc $a \in A$, and the flow-conservation constraints, i.e., the sum of the flow values on the arcs incident to a node $v \neq s, t$ is the same as the sum of the flow values on the arcs incident from v. The value of a flow is $\sum_{a=(v,t)\in A} f(a) - \sum_{a=(t,v)\in A} f(a)$. The decision problem that is proved to be $\mathcal{P}$-complete is that of deciding the parity of the maximum flow value.

There is a collection of $\mathcal{P}$-complete problems that are related to simple polynomial-time algorithms. The *maximal stable set problem*, in which the objective is to find a maximal (not maximum) stable set in an undirected graph, can clearly be solved by a simple greedy algorithm. When using this procedure, we usually select the first available node in each step, and so we find a specific solution, the lexicographically first one. Cook showed that finding the lexicographically first maximal stable set is $\mathcal{P}$-complete. This result might be surprising since this problem is easy to solve in polynomial time. However, $\mathcal{P}$-completeness also provides evidence that the problem is not solvable efficiently in parallel. Consequently, this completeness result supports the intuition that the greedy algorithm is inherently sequential.

4.3. *Parallel computation*

Parallel computation gives us the potential of substantially increasing the size of the instances for which certain problems are manageable by solving them with a large number of processors simultaneously. In studying parallel algorithms, we shall not be concerned with the precise number of parallel processors used, but rather their order as a function of the input size. We say that a parallel algorithm using $\mathrm{O}(p(n))$ processors achieves *optimal speedup*, if it runs in $\mathrm{O}(t(n))$ time and the best sequential algorithm known for solving the same problem runs in $\mathrm{O}(t(n)p(n))$ time. Efficient algorithms that reach optimal (or near-optimal) speedup with a significant number of processors have been found for many of the basic combinatorial problems. Another aspect of parallel computation is concerned with the inherent limitations of using many processors to speed up a computation. To be somewhat realistic, we shall only be interested in algorithms that use a polynomial number of processors. Consequently, we will focus on the possible speedup of polynomial-time sequential computation by parallel processing.

First we define a model of parallel computation. Although many such models have been proposed, one that seems to be the most convenient for designing algorithms is the *parallel random access machine (PRAM).* The PRAM is the parallel analog of the random access machine; it consists of a sequence of random access machines called processors, each with its own infinite local random access memory, in addition to an infinite shared random access memory where each memory cell can store any integer, and the input is stored in the shared memory. Each processor knows the input size and its identification number, although otherwise the processors are identical (i.e., they run the same program). Different variants of the basic PRAM model are distinguished by the manner in which they handle read and write conflicts. In an *EREW PRAM* (exclusive-read, exclusive-write PRAM), for example, it is assumed that each cell of the shared memory is only read from and written into by at most one processor at a time. At the other extreme, in a *CRCW PRAM* (concurrent-read, concurrent-write PRAM), each cell of the memory can be read from and written into by more than one processor at a time. If different processors attempt to write different things in the same cell, then the lowest-numbered processor succeeds.

To illustrate the power of parallel computation, we give parallel algorithms for a problem that we have already discussed. Although finding the lexicographically first maximal stable set is $\mathcal{P}$-complete, Karp and Wigderson have proved, surprisingly, that a maximal stable set can be found efficiently in parallel. Similar, much simpler and more efficient randomized algorithms have subsequently been independently discovered by Luby and by Alon, Babai and Itai.

Consider the most natural sequential algorithm for the following problem: select a node v and include it in the stable set, delete v and all of its neighbors from the graph; repeat this procedure until all nodes have been deleted. Note that this algorithm requires n iterations for a path of length $2n$. A similar approach can still be used for a parallel algorithm. To make the algorithm fast, one selects a stable set in each iteration (rather than a single node), where the set is deleted along with its neighborhood. The following simple way to choose this stable set is due to Luby. A processor is assigned to each node and each edge of the graph. For a graph of order n, the processor assigned to node v picks a random integer $c(v)$ from 1 to n^4. Next, each processor assigned to an edge compares the values at the two nodes of the edge. The stable set selected consists of those nodes v for which $c(v)$ is strictly larger than the values assigned to any of its neighbors.

This algorithm clearly finds a maximal stable set, but it is less clear that few iterations are needed. It can be shown that each iteration is expected to remove a constant fraction of the edges, and consequently, the expected number of iterations is $O(\log n)$. The algorithm can be implemented on a randomized CRCW PRAM in $O(\log n)$ time (if we assume that a processor can choose a random number of size $O(\log n)$ in one step).

Karp and Wigderson introduced a technique that can be used to convert certain randomized algorithms into deterministic ones. The technique can be used if, in the analysis of the randomized algorithm, it is not necessary to assume mutual independence, but, for example, d-wise independent choices suffice for some constant d. One can appeal to known constructions to show that such variables can be chosen from a sample space of polynomial size. Each iteration can then be run for each point in the sample space simultaneously, and this ensures that a good sample point is used. This method can be used to convert the above randomized algorithm into a deterministic one.

When discussing parallel algorithms we shall assume that all arithmetic operations are restricted to polynomial-size numbers, and the number of processors used is polynomially bounded. We define the class $\mathcal{NC}$ to consist of all languages L for which there exists a parallel algorithm that runs in time bounded by a polynomial in $\log n$. Note that in this definition the distinction between the different versions of the basic PRAM model are not relevant. If a problem can be solved by a CRCW PRAM using $p(n)$ processors in $O(\log^i n)$ time, then it can be solved by an EREW PRAM using $p(n)$ processors and $O(\log^i n \log p(n))$ time. The maximal stable set algorithm of Luby discussed earlier uses a randomized version of the CRCW PRAM. We define the complexity class $\mathcal{RNC}$ to be the $\mathcal{NC}$ analog of $\mathcal{RP}$.

It is straightforward to see that the Boolean product of two $n \times n$ (0,1)-matrices can be computed in constant time on a CRCW PRAM using $O(n^3)$ processors. By repeatedly squaring the adjacency matrix of a graph, the directed reachability problem can be solved in $O(\log n)$ time. This is, in some sense, the generic problem in $\mathcal{NL}$, and more generally, any problem in $\mathcal{NL}$ can be solved by a CRCW PRAM in $O(\log n)$ time. As a consequence, a log-space reduction can be simulated efficiently in parallel, and therefore $\mathcal{P}$-completeness provides evidence that a problem is not efficiently solvable in parallel.

Theorem 4.8. *For any $\mathcal{P}$-complete problem L, $L \in \mathcal{NC}$ if and only if $\mathcal{NC} = \mathcal{P}$.*

We get the following chain of inclusions:

$$\mathcal{L} \subseteq \mathcal{NL} \subseteq \mathcal{NC} \subseteq \mathcal{P} \subseteq \mathcal{NP} \subseteq \mathcal{PSPACE}.$$

On the other hand, the computation of a CRCW PRAM that runs in $O(\log^i n)$ time can be simulated by a Turing machine in $O(\log^{i+1} n)$ space. This proves that $\mathcal{NC}$ is contained in $\bigcup_{i \geqslant 1} \mathrm{DSPACE}(\log^i n)$. By the analog of Theorem 2.12 for space complexity, this implies that $\mathcal{NC} \neq \mathcal{PSPACE}$.

We have already seen that a simple parallel algorithm for the directed reachability problem is based on matrix multiplication, and in fact, many simple parallel graph algorithms are based on matrix operations. Csanky has given an $\mathcal{NC}$ algorithm to compute the rank and the determinant of a matrix over the reals in $O(\log^2 n)$ time. As a corollary, we get a parallel algorithm to solve systems of linear equations. Berkowitz, Chistov and Mulmuley have extended these results to matrices over arbitrary fields. One of the most beautiful connections between matrix operations and graph algorithms is that a perfect matching in a graph can be found by an efficient randomized parallel algorithm using only a single matrix inversion (see chapter 3).

There has been substantial work over the past several years in finding efficient parallel algorithms for combinatorial problems. Some of these algorithms are mentioned elsewhere in this Handbook. For further results and more details the interested reader is referred to the survey of Karp and Ramachandran (1990).

Acknowledgements

In writing this survey, we have often been aided by the expertise of others. In particular, we would like to thank Laci Babai and Michael Sipser for sharing with us many of their insights into complexity theory. The surveys that we have cited in particular areas of complexity theory were extremely useful and we wish express our gratitude to their authors. In addition, we would like to thank Shafi Goldwasser, Leslie Hall, David Johnson, Jan Karel Lenstra, Laci Lovász, Albert Meyer, Carolyn Haibt Norton and Larry Stockmeyer for their many insightful comments on an earlier draft. The research of the first author was supported in part by an NSF Presidential Young Investigator award CCR-89-96272 with matching support from IBM, UPS and Sun and by Air Force Contract AFOSR-86-0078. The research of the second author was supported in part by Air Force Contract AFOSR-86-0078 and by a Packard Research Fellowship. Both were supported in part by the National Science Foundation, the Air Force Office of Scientific Research, and the Office of Naval Research, through NSF grant DMS-8920550.

References

Aho, A.V., J.E. Hopcroft and J.D. Ullman

[1974] *The Design and Analysis of Computer Algorithms* (Addison-Wesley, Reading, MA).

Aleliunas, R., R.M. Karp, R.J. Lipton, L. Lovász and C. Rackoff

[1979] Random walks, universal traversal sequences, and the complexity of maze problems, in: *Proc. 20th Annu. IEEE Symp. on Foundations of Computer Science* (IEEE Computer Society Press, Washington, DC) pp. 218–223.

Babai, L.

[1985] Trading group theory for randomness, in: *Proc. 17th Annu. ACM Symp. on Theory of Computing* (ACM Press, New York) pp. 421–429.

Baker, T., J. Gill and R. Solovay

[1975] Relativizations of the $\mathscr{P} = ?\mathscr{NP}$ question, *SIAM J. Comput.* **4**, 431–442.

Chandra, A.K., D.C. Kozen and L.J. Stockmeyer

[1981] Alternation, *J. Assoc. Comput. Mach.* **28**, 114–133.

Cobham, A.

[1965] The intrinsic computational difficulty of functions, in: *Proc. 1964 Int. Congr. for Logic, Methodology and Philosophy of Science*, ed. Y. Bar-Hillel (North-Holland, Amsterdam) pp. 24–30.

Coffman Jr, E.G., G.S. Lueker and A.H.G. Rinnooy Kan

[1988] Asymptotic methods in the probabilistic analysis of sequencing and packing heuristics, *Management Sci.* **34**, 266–290.

Cook, S.A.

[1971] The complexity of theorem-proving procedures, in: *Proc. 3rd Annu. ACM Symp. on Theory of Computing* (ACM Press, New York) pp. 151–158.

Davis, M.

[1977] Unsolvable problems, in: *Handbook of Mathematical Logic*, ed. J. Barwise (North-Holland, Amsterdam) pp. 567–594.

Edmonds, J.

[1965] Paths, trees, and flowers, *Canad. J. Math.* **17**, 449–467.

Fischer, M.J., and M.O. Rabin

[1974] Super-exponential complexity of Presburger arithmetic, in: *Complexity of Computation, SIAM–AMS Proc.*, Vol. 7, ed. R.M. Karp (American Mathematical Society, Providence, RI) pp. 27–41.

Furst, M., J.B. Saxe and M. Sipser

[1981] Parity, circuits, and the polynomial-time hierarchy, *Math. Systems Theory* **17**, 13–27.

Garey, M.R., and D.S. Johnson

[1976] The complexity of near-optimal graph coloring, *J. Assoc. Comput. Mach.* **23**, 43–49.

[1979] *Computers and Intractability: A Guide to the Theory of NP-Completeness* (W.H. Freeman, San Francisco, CA).

Gödel, K.

[1931] Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, I, *Monatsh. Math. Phys.* **38**, 173–198.

Goldwasser, S.

[1989] Interactive proof systems, in: *Computational Complexity Theory, AMS Symposia in Applied Mathematics*, Vol. 38, ed. J. Hartmanis (American Mathematical Society, Providence, RI) pp. 108–128.

Goldwasser, S., S. Micali and C. Rackoff

[1989] The knowledge complexity of interactive proof-systems, *SIAM J. Comput.* **18**, 186–208.

Graham, R.L.

[1966] Bounds for certain multiprocessing anomalies, *Bell System Tech. J.* **45**, 1563–1581.

Hartmanis, J., and R.E. Stearns

[1965] On the computational complexity of algorithms, *Trans. Amer. Math. Soc.* **117**, 285–306.

Hopcroft, J., W. Paul and L. Valiant

[1977] On time versus space, *J. Assoc. Comput. Mach.* **24**, 332–337.

Hopcroft, J.E., and J.D. Ullman

[1979] *Introduction to Automata Theory, Languages, and Computation* (Addison-Wesley, Reading, MA).

Immerman, N.

[1988] Nondeterministic space is closed under complementation, *SIAM J. Comput.* 17, 935–938.

Johnson, D.S.

[1984] The NP-completeness column: an ongoing guide, *J. Algorithms* 5, 284–299.

[1988] The NP-completeness column: an ongoing guide, *J. Algorithms* **9**, 426–444.

[1992] The NP-completeness column: an ongoing guide, *J. Algorithms* **13**, 502–524.

Karp, R.M.

[1972] Reducibility among combinatorial problems, in: *Complexity of Computer Computations*, eds. R.E. Miller and J.W. Thatcher (Plenum Press, New York) pp. 85–103.

[1976] The probabilistic analysis of some combinatorial search algorithms, in: *Algorithms and Complexity: New Directions and Recent Results*, ed. J.F. Traub (Academic Press, New York) pp. 1–19.

Karp, R.M., and V. Ramachandran

[1990] Parallel algorithms for shared-memory machines, in: *Algorithms and Complexity, Handbook of Theoretical Computer Science*, Vol. A, ed. J. van Leeuwen (Elsevier, Amsterdam) pp. 869–941.

Karp, R.M., J.K. Lenstra, C.J.H. McDiarmid and A.H.G. Rinnooy Kan

[1985] Probabilistic analysis, in: *Combinatorial Optimization: Annotated Bibliographies*, eds. M. O'hEigeartaigh, J.K. Lenstra and A.H.G. Rinnooy Kan (Centre for Mathematics and Computer Science/Wiley, Amsterdam/Chichester) pp. 52–88.

Lenstra, A.K., and H.W. Lenstra Jr

[1990] Algorithms in number theory, in: *Algorithms and Complexity, Handbook of Theoretical Computer Science*, Vol. A, ed. J. van Leeuwen (Elsevier, Amsterdam) pp. 673–715.

Levin, L.A.

[1973] Universal sequential search problems, *Problemy Peredachi Informatsii* 9, 115–116 [*Problems Inform. Transmission* 9, 265–266].

[1986] Average case complete problems, *SIAM J. Comput.* **15**, 285–286.

Meyer, A.R., and L.J. Stockmeyer

[1972] The equivalence problem for regular expressions with squaring requires exponential time, in: *Proc. 13th Annu. Symp. on Switching and Automata Theory* (IEEE Computer Society Press, Washington, DC) pp. 125–129.

Paul, W.J., N. Pippenger, E. Szemerédi and W.T. Trotter

[1983] On determinism versus non-determinism and related problems, in: *Proc. 24th IEEE Symp. on Foundations of Computer Science* (IEEE Computer Society Press, Washington, DC) pp. 429–438.

Rabin, M.O.

[1976] Probabilistic algorithms, in: *Algorithms and Complexity*, ed. J.F. Traub (Academic Press, New York) pp. 21–40.

Savitch, W.J.

[1970] Relationships between nondeterministic and deterministic tape complexities, *J. Comput. System Sci.* **4**, 177–192.

Shamir, R.

[1987] The efficiency of the simplex method: a survey, *Management Sci.* **33**, 301–334.

Shmoys, D.B.

[1995] Computing near-optimal solutions to combinatorial optimization problems, in: *Combinatorial Optimization*, eds. W. Cook, L. Lovász and P. Seymour, *DIMACS Series in Discrete Mathematics and Theoretical Computer Science* (AMS, Providence, RI) to appear.

Solovay, R., and V. Strassen

[1977] A fast Monte-Carlo test for primality, *SIAM J. Comput.* **6**, 84–85.

Stockmeyer, L.J.

[1987] Classifying the computational complexity of problems, *J. Symbolic Logic* **52**, 1–43.

Szelepcsényi, R.

[1987] The method of forcing for nondeterministic automata, *Bull. European Assoc. Theor. Comput. Sci.* **33**, 96–100.

Turing, A.M.

[1936] On computable numbers, with an application to the Entscheidungsproblem, *Proc. London Math. Soc. II* **2**, 230–265.

Valiant, L.G.

[1979] The complexity of computing the permanent, *Theor. Comput. Sci.* **8**, 189–201.

van Leeuwen, J.

[1990] *Handbook of Theoretical Computer Science* (Elsevier, Amsterdam).

Yao, A.C.-C.

[1985] Separating the polynomial-time hierarchy by oracles, in: *Proc. 26th Annu. IEEE Symp. on Foundations of Computer Science* (IEEE Computer Society Press, Washington, DC) pp. 1–10.

CHAPTER 30

Polyhedral Combinatorics

Alexander SCHRIJVER

CWI, Kruislaan 413, 1098 SJ Amsterdam, The Netherlands

Contents

HANDBOOK OF COMBINATORICS
Edited by R. Graham, M. Grötschel and L. Lovász

1. Introduction

Polyhedral combinatorics studies combinatorial problems with the help of polyhedra. Let us first give a simple, illustrative example. Let $G = (V, E)$ be a graph, and let $c\colon E \to \mathbb{R}_+$ be a weight function on the edges of G. Suppose we want to find a matching M in G with "weight"

$$c(M) = \sum_{e \in M} c(e) \tag{1.1}$$

as large as possible, Thus we want to "solve"

$$\max\{c(M) \mid M \text{ matching in } G\} . \tag{1.2}$$

Denote for any matching M, the incidence vector of M in $\mathbb{R}^E$ by χ^M, i.e., $\chi^M(e) := 1$ if $e \in M$ and $:= 0$ if $e \not\in M$. Considering the weight function $c\colon E \to \mathbb{R}$ as a *vector* in $\mathbb{R}^E$, we can write problem (1.2) as

$$\max\{c^{\mathrm{T}}\chi^M \mid M \text{ matching in } G\} . \tag{1.3}$$

This amounts to maximizing a linear function over a finite set of vectors. Hence we can equally well maximize over the *convex hull* of these vectors:

$$\max\{c^{\mathrm{T}}x \mid x \in \operatorname{conv}\{\chi^M \mid M \text{ matching in } G\}\} . \tag{1.4}$$

The set

$$\operatorname{conv}\{\chi^M \mid M \text{ matching in } G\} \tag{1.5}$$

is a polytope in $\mathbb{R}^E_+$, called the *matching polytope* of G. It follows that there exist a matrix A and a vector b such that

$$\operatorname{conv}\{\chi^M \mid M \text{ matching in } G\} = \{x \in \mathbb{R}^E \mid x \geqslant 0, Ax \leqslant b\} . \tag{1.6}$$

Then problem (1.4) is equal to

$$\max\{c^{\mathrm{T}}x \mid x \geqslant 0, Ax \leqslant b\} . \tag{1.7}$$

In this way we have formulated the original combinatorial problem (1.2) as a *linear programming problem*. This enables us to apply linear programming methods to study the original problem.

The problem at this point is, however, how to find the matrix A and the vector b. We know that A and b exist, but we must know them in order to apply linear programming methods.

If G is bipartite, it turns out that the matching polytope of G is equal to the set of all vectors $x \in \mathbb{R}^E$ satisfying

$$\begin{aligned} &x(e) \geqslant 0 , \quad e \in E \\ &\sum_{e \ni v} x(e) \leqslant 1 , \quad v \in V . \end{aligned} \tag{1.8}$$

That is, for A we can take the $V \times E$ incidence matrix of G and for b the all-one vector $\mathbf{1}$ in $\mathbb{R}^V$.

It is not difficult to show that the matching polytope for bipartite graphs is indeed completely determined by (1.8). First note that the matching polytope is contained in the polytope defined by (1.8), since χ^M satisfies (1.8) for each matching M. To see the converse inclusion, we note that if G is bipartite, then the matrix A is *totally unimodular*, i.e., each square submatrix of A has determinant belonging to $\{0, +1, -1\}$. This may be seen to imply that the vertices of the polytope determined by (1.8) are *integral* vectors, i.e., they belong to $\mathbb{Z}^E$. Now each integral vector satisfying (1.8) must trivially be equal to χ^M for some matching M. Hence the polytope determined by (1.8) is equal to the matching polytope of G.

For each *nonbipartite* graph, the matching polytope is not completely determined by (1.8). Indeed, if C is an odd circuit in G, then the vector $x \in \mathbb{R}^E$ defined by $x(e) = \frac{1}{2}$ if $e \in C$ and 0 if $e \not\in C$, satisfies (1.8) but does not belong to the matching polytope.

In fact, it is a pioneering theorem in polyhedral combinatorics due to J. Edmonds that gives a complete description of the inequalities needed to describe the matching polytope for arbitrary graphs.

When we have formulated the matching problem as LP problem (1.7), we can apply LP techniques to study the matching problem. Thus we can find a maximum weighted matching in a bipartite graph, e.g., with the simplex method. Moreover, the Duality Theorem of Linear Programming gives

$$\begin{aligned} &\max\{c(M) \mid M \text{ matching in } G\} = \max\{c^{\mathrm{T}}x \mid x \geq 0, Ax \leq \mathbf{1}\} \\ &\quad = \min\{y^{\mathrm{T}}\mathbf{1} \mid y \geq 0, y^{\mathrm{T}}A \geq c^{\mathrm{T}}\} . \qquad (1.9) \end{aligned}$$

In the special case of G bipartite and c being the all-one vector in $\mathbb{R}^E$, we can derive from this the *König–Egerváry Theorem*. The left-most expression in (1.9) is equal to the maximum size of a matching. The minimum can be seen to be attained by an integral vector y, again by the total unimodularity of A. This y is a $\{0, 1\}$-vector in $\mathbb{R}^V$, and is the incidence vector of some subset W of V intersecting every edge of G. Thus (1.9) implies that the maximum size of a matching is equal to the minimum size of a set of vertices intersecting all edges of G.

As an extension, one can derive the *Tutte–Berge Formula* from the inequality system given by Edmonds for arbitrary graphs.

Bipartite matching forms an easy example in polyhedral combinatorics. We now discuss the central idea of polyhedral combinatorics – taking convex hulls – in a more general framework.

Let $\mathcal{F}$ be a collection of subsets of a finite set S, let $c: S \to \mathbb{R}$, and suppose we are interested in

$$\max\left\{\sum_{s \in U} c(s) \mid U \in \mathcal{F}\right\} . \qquad (1.10)$$

(For example, S is the set of edges of a graph, and $\mathcal{F}$ is the collection of matchings, in which case (1.10) is the maximum "weight" of a matching.) Usually, $\mathscr{F}$ is too large to evaluate each set U in $\mathscr{F}$ in order to determine the maximum (1.10). (For example, the collection of matchings is exponentially large in the size of the graph.) Now (1.10) is equal to

$$\max\{c^{\mathrm{T}}\chi^{U} \mid U \in \mathscr{F}\}\,, \tag{1.11}$$

where χ^U denotes the *incidence vector* of U in $\mathbb{R}^S$, i.e., $\chi^U(s) = 1$ if $s \in U$ and 0 otherwise. [Here we identify functions $c\colon S \to \mathbb{R}$ with vectors in the linear space $\mathbb{R}^S$, and accordingly we shall sometimes denote $c(s)$ by c_s.] Since (1.11) means maximizing a linear function over a finite set of vectors, we could equally well maximize over the convex hull of these vectors:

$$\max\{c^{\mathrm{T}}x \mid x \in \operatorname{conv}\{\chi^U \mid U \in \mathscr{F}\}\}\,. \tag{1.12}$$

Since this convex hull is a polytope, there exist a matrix A and a column vector b such that

$$\operatorname{conv}\{\chi^U \mid U \in \mathscr{F}\} = \{x \in \mathbb{R}^S \mid Ax \leqslant b\}\,. \tag{1.13}$$

Hence (1.12) is equal to

$$\max\{c^{\mathrm{T}}x \mid Ax \leqslant b\}\,. \tag{1.14}$$

Thus we have formulated the original combinatorial problem as a linear programming problem, and we can appeal to linear programming methods to study the combinatorial problem.

In order to determine the maximum (1.10) algorithmically, we could use LP algorithms like the simplex method or the primal–dual method. Sometimes, with the ellipsoid method the polynomial-time solvability of (1.10) can be shown. Moreover, by the Duality Theorem of Linear Programming, problem (1.14), and hence problem (1.10), is equal to

$$\min\{y^{\mathrm{T}}b \mid y \geqslant 0,\ y^{\mathrm{T}}A = c^{\mathrm{T}}\}\,, \tag{1.15}$$

which gives us a min–max equation for the combinatorial maximum. Often this provides us with a "good characterization" [i.e., problem (1.10) belongs to NP ∩ co-NP], and it enables us to carry out a "sensitivity analysis" of the combinatorial problem, etc.

However, in order to apply LP techniques, we should be able to find matrix A and vector b satisfying (1.13). This is one of the main theoretical problems in polyhedral combinatorics.

Often, one first "guesses" a system $Ax \leqslant b$, and next, one tries to prove that $Ax \leqslant b$ forms a complete description of the polytope. Sometimes, like in bipartite matching, this can be shown with the help of the total unimodularity of A. However, in general A is not totally unimodular, and one has to try more complicated techniques to show that $Ax \leqslant b$ completely describes the polytope. In this survey, we mention the techniques of "total dual integrality", "blocking polyhedra", "anti-blocking polyhedra", and "cutting planes".

In several cases, the guessed system $Ax \leqslant b$ turns out not to be a complete description, but just gives an approximation of the polytope. This can still be useful, since in that case the linear programming problem $\max\{c^{\mathrm{T}}x \mid Ax \leqslant b\}$ gives a (hopefully good) upper bound for the combinatorial maximum. This can be very useful in a so-called *branch-and-bound* algorithm for the combinatorial problem.

Historically, applying LP techniques to combinatorial problems came along with the introduction of linear programming in the 1940s and 1950s. Dantzig, Ford, Fulkerson, Hoffman, Johnson and Kruskal studied problems like the transportation, flow, and assignment problems, which can be reduced to linear programming (by the total unimodularity of the constraint matrix), and the traveling salesman problem, using a rudimentary version of a cutting plane technique (extended by Gomory to general integer linear programming).

The field of polyhedral combinatorics was extended and deepened considerably by the work of Edmonds in the 1960s and 1970s. He characterized basic polytopes like the matching polytope, the arborescence polytope, and the matroid intersection polytope; he introduced (with Giles) the important concept of total dual integrality; and he advocated the link between polyhedra, min–max relations, good characterizations, and polynomial-time solvability. Fulkerson designed the clarifying framework of blocking and anti-blocking polyhedra, enabling the deduction of one polyhedral characterization or min–max relation from another.

In this chapter we describe the basic techniques in polyhedral combinatorics, and we derive as illustrations polyhedral characterizations for some concrete combinatorial problems. First, in sections 2 and 3, we give some background information on polyhedra and linear programming methods.

For background and related literature we refer to Grötschel et al. (1988), Grötschel and Padberg (1985), Grünbaum (1967), Lovász (1977, 1979), Pulleyblank (1983), Schrijver (1983b, 1986), and Stoer and Witzgall (1970).

2. Background information on polyhedra

For an in-depth survey on polyhedra (focusing on the combinatorial properties) we refer the reader to chapter 18. In this section, we give a brief review on polyhedra, covering those parts of polyhedral theory required for the present chapter.

A set $P \subseteq \mathbb{R}^n$ is called a *polyhedron* if there exist a matrix A and a column vector b such that

$$P = \{x \mid Ax \leqslant b\} . \tag{2.1}$$

If (2.1) holds, we say that $Ax \leqslant b$ *determines* P. A set $P \subseteq \mathbb{R}^n$ is called a *polytope* if there exist $x_1, \ldots, x_t$ in $\mathbb{R}^n$ such that $P = \mathrm{conv}\{x_1, \ldots, x_t\}$. The following theorem is intuitively clear, but is not trivial to prove, and is usually attributed to Minkowski (1896), Steinitz (1916), and Weyl (1935).

Finite Basis Theorem for Polytopes 2.2. *A set P is a polytope if and only if P is a bounded polyhedron.*

Motzkin, in 1936, extended this to:

Decomposition Theorem for Polyhedra 2.3. *$P \subseteq \mathbb{R}^n$ is a polyhedron if and only if there exist $x_1, \ldots, x_t, y_1, \ldots, y_s \in \mathbb{R}^n$ such that*

$$P = \{\lambda_1 x_1 + \cdots + \lambda_t x_t + \mu_1 y_1 + \cdots + \mu_s y_s \mid \lambda_1, \ldots, \lambda_t, \mu_1, \ldots, \mu_s \geq 0; \\ \lambda_1 + \cdots + \lambda_t = 1\} .$$

Now let $P = \{x \mid Ax \leq b\}$ be a nonempty polyhedron, where A has order $m \times n$. If $c \in \mathbb{R}^n$ with $c \neq 0$ and $\delta = \max\{c^{\mathrm{T}}x \mid x \in P\}$, then the set $\{x \mid c^{\mathrm{T}}x = \delta\}$ is called a *supporting hyperplane* of P. A subset F of P is called a *face* of P if $F = P$ or if $F = P \cap H$ for some supporting hyperplane H of P. Clearly, a face of P is a polyhedron again. It can be shown that for any face F of P there exists a subsystem $A'x \leq b'$ of $Ax \leq b$ such that $F = \{x \in P \mid A'x = b'\}$. Hence P has only finitely many faces. They are ordered by inclusion. *Minimal faces* are the faces minimal with respect to inclusion. The following theorem is due to Hoffman and Kruskal (1956).

Theorem 2.4. *A set F is a minimal face of P if and only if $\emptyset \neq F \subseteq P$ and*

$$F = \{x \mid A'x = b'\}$$

for some subsystem $A'x \leq b'$ of $Ax \leq b$.

All minimal faces have the same dimension, viz. n-rank(A). If this is 0, minimal faces correspond to vertices: a *vertex* of P is an element of P which is not a convex combination of two other elements of P. Only if rank$(A) = n$, does P have vertices, and then those vertices are exactly the minimal faces. Hence:

Theorem 2.5. *Vector z in P is a vertex of P if and only if $A'z = b'$ for some subsystem $A'x \leq b'$ of $Ax \leq b$, with A' nonsingular of order n.*

The matrix A' (or subsystem $A'x \leq b'$) is sometimes called a *basis* for z. Generally, such a basis is not unique. P is called *pointed* if it has vertices. A polytope is always pointed, and is the convex hull of its vertices.

Two vertices x and y of P are *adjacent* if conv$\{x, y\}$ is a face of P. It can be shown that if P is a polytope, then two vertices x and y are adjacent if and only if the vector $\frac{1}{2}(x + y)$ is not a convex combination of other vertices of P. Moreover, one can show:

Theorem 2.6. *Vertices z' and z'' of the polyhedron P are adjacent if and only if z' and z'' have bases $A'x \leqslant b'$ and $A''x \leqslant b''$, respectively, so that they have exactly $n-1$ constraints in common.*

The polyhedron P gives rise to a graph, whose nodes are the vertices of P, two of them being adjacent in the graph if and only if they are adjacent on P. The *diameter* of P is the diameter of this graph. The following conjecture is due to W. M. Hirsch (cf. Dantzig 1963).

Hirsch's Conjecture 2.7. A polytope in $\mathbb{R}^n$ determined by m inequalities has diameter at most $m-n$.

This conjecture is related to the number of iterations in the simplex method (see section 3). See also Klee and Walkup (1967) and Larman (1970). [The Hirsch conjecture was proved by Naddef (1989) for polytopes all of whose vertices are $\{0,1\}$ – vectors.]

A *facet* of P is an inclusion-wise maximal face F of P with $F \neq P$. A face F of P is a facet if and only if $\dim(F) = \dim(P) - 1$. An inequality $c^{\mathrm{T}}x \leqslant \delta$ is called a *facet-inducing inequality* if $P \subseteq \{x \mid c^{\mathrm{T}}x \leqslant \delta\}$ and $P \cap \{x \mid c^{\mathrm{T}}x = \delta\}$ is a facet of P.

Suppose $Ax \leqslant b$ is an *irredundant* (or *minimal*) system determining P, i.e., no inequality in $Ax \leqslant b$ is implied by the other. Let $A^+x \leqslant b^+$ be those inequalities $a^{\mathrm{T}}x \leqslant \beta$ from $Ax \leqslant b$ for which $a^{\mathrm{T}}z < \beta$ for at least one z in P. Then each inequality in $A^+x \leqslant b^+$ is a facet-inducing inequality. Moreover, this defines a one-to-one relation between facets and inequalities in $A^+x \leqslant b^+$. If P is full-dimensional, then the irredundant system $Ax \leqslant b$ is unique up to multiplication of inequalities by positive scalars. The following characterization holds.

Theorem 2.8. *If $P = \{x \mid Ax \leqslant b\}$ is full-dimensional, then $Ax \leqslant b$ is irredundant if and only if for each pair $a_i^{\mathrm{T}}x \leqslant b_i$ and $a_j^{\mathrm{T}}x \leqslant b_j$ of constraints from $Ax \leqslant b$ there is a vector x' in P satisfying $a_i^{\mathrm{T}}x' = b_i$ and $a_j^{\mathrm{T}}x'' < b_j$.*

The polyhedron P is called *rational* if we can take A and b in (2.1) rational-valued (and hence we can take them integer-valued). P is rational if and only if the vectors $x_1, \ldots, x_t$ and $y_1, \ldots, y_s$ in Theorem 2.3 can be taken to be rational. P is called *integral* if we can take $x_1, \ldots, x_t$ and $y_1, \ldots, y_s$ in Theorem 2.3 integer-valued. Hence P is integral if and only if P is the convex hull of the integer vectors in P or, equivalently, if and only if every minimal face of P contains integer vectors.

3. Background information on linear programming

Linear programming, abbreviated by LP, studies the problem of maximizing or minimizing a linear function $c^{\mathrm{T}}x$ over a polyhedron P. Examples of such a problem are:

$$\begin{aligned} &\text{(i)} && \max\{c^{\mathrm{T}}x \mid Ax \leq b\}\,, \\ &\text{(ii)} && \max\{c^{\mathrm{T}}x \mid x \geq 0,\, Ax \leq b\}\,, \\ &\text{(iii)} && \max\{c^{\mathrm{T}}x \mid x \geq 0,\, Ax = b\}\,, \\ &\text{(iv)} && \min\{c^{\mathrm{T}}x \mid x \geq 0,\, Ax \geq b\}\,. \end{aligned} \tag{3.1}$$

It can be shown, for each of the problems (i)–(iv), that if the set involved is a polyhedron with vertices [for (ii)–(iv) this follows if it is nonempty], and if the optimum value is finite, then it is attained by a vertex of the polyhedron.

Each of the optima (3.1) is equal to the optimum value in some other LP problem, called the *dual problem*.

Duality Theorem of Linear Programming 3.2. *Let A be an $m \times n$ matrix and let $b \in \mathbb{R}^m$ and $c \in \mathbb{R}^n$. Then*

$$\begin{aligned} &\text{(i)} && \max\{c^{\mathrm{T}}x \mid Ax \leq b\} &&= \min\{y^{\mathrm{T}}b \mid y \geq 0,\, y^{\mathrm{T}}A = c^{\mathrm{T}}\}\,; \\ &\text{(ii)} && \max\{c^{\mathrm{T}}x \mid x \geq 0,\, Ax \leq b\} &&= \min\{y^{\mathrm{T}}b \mid y \geq 0,\, y^{\mathrm{T}}A \geq c^{\mathrm{T}}\}\,; \\ &\text{(iii)} && \max\{c^{\mathrm{T}}x \mid x \geq 0,\, Ax = b\} &&= \min\{y^{\mathrm{T}}b \mid y^{\mathrm{T}}A \geq c^{\mathrm{T}}\}\,; \\ &\text{(iv)} && \min\{c^{\mathrm{T}}x \mid x \geq 0,\, Ax \geq b\} &&= \max\{y^{\mathrm{T}}b \mid y \geq 0,\, y^{\mathrm{T}}A \leq c^{\mathrm{T}}\}\,; \end{aligned} \tag{3.3}$$

provided that these sets are nonempty.

It is not difficult to derive this from:

Farkas's Lemma 3.4. *Let A be an $m \times n$ matrix and let $b \in \mathbb{R}^m$. Then $Ax = b$ has a solution $x \geq 0$ if and only if $y^{\mathrm{T}}b \geq 0$ holds for each vector $y \in \mathbb{R}^m$ with $y^{\mathrm{T}}A \geq 0$.*

The principle of *complementary slackness* says: let x and y satisfy $Ax \leq b$, $y \geq 0$, $y^{\mathrm{T}}A = c^{\mathrm{T}}$ then x and y are optimum solutions in Theorem 3.2(i) if and only if $y_i = 0$ or $a_i^{\mathrm{T}}x = b_i$ for each $i = 1, \ldots, m$ (where $a_i^{\mathrm{T}}x = b_i$ denotes the ith line in the system $Ax = b$). Similar statements hold for Theorem 3.2(ii)–(iv).

We now describe briefly three of the methods for solving LP problems. The first two methods, the famous simplex method and the primal–dual method, can be considered also, when applied to combinatorial problems, as a guideline to deriving a "combinatorial" algorithm from a polyhedral characterization. The third method, the ellipsoid method, is more of theoretical value: it is a tool sometimes used to derive the polynomial-time solvability of a combinatorial problem.

3.1. The simplex method

The simplex method, due to Dantzig (1951a), is the method used most often for linear programming. Let $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, and $c \in \mathbb{R}^n$. Suppose we wish to solve $\max\{c^T x \mid Ax \leq b\}$, where the polyhedron $P := \{x \mid Ax \leq b\}$ is a polyhedron with vertices, i.e., $\text{rank}(A) = n$.

The idea of the simplex method is to make a trip, going from a vertex to a better adjacent vertex, until an optimal vertex is reached. By Theorem 2.5, vertices can be described by bases, while by Theorem 2.6 adjacency can be described by bases differing in exactly one constraint. Thus the process can be described by a series

$$A_0 x \leq b_0, A_1 x \leq b_1, A_2 x \leq b_2, \ldots \tag{3.5}$$

of bases, where each $x_k := A_k^{-1} b_k$ is a vertex of P, where $A_{k+1} x \leq b_{k+1}$ differs by one constraint from $A_k x \leq b_k$, and where $c^T x_{k+1} \geq c^T x_k$.

The series can be found as follows. Suppose $A_k x \leq b_k$ has been found. If $c^T A_k^{-1} \geq 0$, then x_k is an optimal solution of max $\{c^T x \mid Ax \leq b\}$, since for each x satisfying $Ax \leq b$ one has $A_k x \leq b_k$ and hence $c^T x = (c^T A_k^{-1}) A_k x \leq (c^T A_k^{-1}) b_k = c^T x_k$.

If $c^T A_k^{-1} \not\geq 0$, choose an index i so that $(c^T A_k^{-1})_i < 0$, and let $z := -A_k^{-1} e_i$ (where e_i denotes the ith unit basis vector in $\mathbb{R}^n$). Note that for $\lambda \geq 0$, $x_k + \lambda z$ traverses an edge or ray of P (i.e., face of dimension 1), or it is outside of P for all $\lambda > 0$. Moreover, $c^T z = -c^T A_k^{-1} e_i > 0$. Now if $Az \leq 0$, then $x_k + \lambda z \in P$ for all $\lambda \geq 0$, whence $\max\{c^T x \mid Ax \leq b\} = \infty$. If $Az \not\leq 0$, let λ_0 be the largest λ such that $x_k + \lambda z$ belongs to P, i.e.,

$$\lambda_0 := \min\left\{\frac{b_j - a_j^T x_k}{a_j^T z} \;\middle|\; j = 1, \ldots, m, a_j^T z > 0\right\}. \tag{3.6}$$

Choose an index j attaining this minimum. Replacing the ith inequality in $A_k x \leq b_k$ by inequality $a_j^T x \leq b_j$ then gives us the next system $A_{k+1} x \leq b_{k+1}$.

Note that $x_{k+1} = x_k + \lambda_0 z$, implying that if $x_{k+1} \neq x_k$ then $c^T x_{k+1} > c^T x_k$. Clearly, the above process stops if $c^T x_{k+1} > c^T x_k$ for each k (since P has only finitely many vertices). This is the case if each vertex has exactly one basis – the *nondegenerate* case. However, in general it can happen that $x_{k+1} = x_k$ for certain k. Several "pivot selection rules", prescribing the choice of i and j above, have been found which could be proved to yield termination of the simplex method. No one of these rules could be proved to give a polynomial-time method – in fact, most of them could be shown to require an exponential number of iterations in the worst case.

The number of iterations in the simplex method is related to the diameter of the underlying polyhedron P. Suppose P is a polytope. If there is a pivot selection rule such that for each $c \in \mathbb{R}^n$ the problem $\max\{c^T x \mid Ax \leq b\}$ can be solved within t iterations of the simplex method (starting with an arbitrary first basis $A_0 x \leq b_0$ corresponding to a vertex), then clearly P has diameter at most t. However, as Padberg and Rao (1974) showed, the "traveling-salesman polytopes" (see section 10) form a class of polytopes of diameter at most 2, while maximizing a linear function over these polytopes is NP-complete.

A main problem seems that we do not have a better criterion for adjacency than Theorem 2.6. Note that a vertex of P can be adjacent to an exponential number of vertices (in the sizes of A and b), whereas for any basis A' there are at most $n(m-n)$ bases differing from A' in exactly one row. In the degenerate case, there can be several bases corresponding to one and the same vertex. Just this phenomenon shows up frequently in polytopes occurring in combinatorial optimization, and one of the main objectives is to find pivoting rules preventing us going through many bases corresponding to the same vertex (cf. Cunningham 1979).

3.2. *Primal–dual method*

As a generalization of similar methods for network flow and transportation problems, Dantzig et al. (1956) designed the "primal–dual method" for LP. The general idea is as follows. Starting with a dual feasible solution y, the method searches for a primal feasible solution x satisfying the complementary slackness condition with respect to y. If such a primal feasible solution is found, x and y form a pair of optimal (primal and dual) solutions. If no such primal solution is found, the method prescribes a modification of y, after which we start anew.

The problem now is how to find a primal feasible solution x satisfying the complementary slackness condition, and how to modify the dual solution y if no such primal solution is found. For general LP problems this problem can be seen to amount to another LP problem, generally simpler than the original LP problem. To solve the simpler problem we could use any LP method, e.g., the simplex method. In many combinatorial applications, however, this simpler LP problem is a simpler combinatorial optimization problem, for which direct methods are available (see Papadimitriou and Steiglitz 1982). Thus, if we can describe a combinatorial optimization problem as a linear program, the primal–dual method gives us a scheme for reducing one combinatorial problem to an easier combinatorial problem.

We shall now describe the primal–dual method more precisely. Suppose we wish to solve the LP problem.

$$\min\{c^{\mathrm{T}}x \mid x \geqslant 0,\ Ax = b\}\ , \tag{3.7}$$

where A is an $m \times n$ matrix, with columns $a_1, \ldots, a_n$, $b \in \mathbb{R}^m$, and $c \in \mathbb{R}^n$. The dual problem is

$$\max\{y^{\mathrm{T}}b \mid y^{\mathrm{T}}A \leqslant c^{\mathrm{T}}\}\ . \tag{3.8}$$

The primal–dual method consists of repeating the following *primal–dual iteration*. Suppose we have a feasible solution y_0 for problem (3.8). Let A' be the submatrix of A consisting of those columns a_j of A for which $y_0^{\mathrm{T}}a_j = c_j$. To find a feasible primal solution for which the complementary slackness condition holds, solve the *restricted linear program*

$$\min\{\lambda \mid x',\ \lambda \geqslant 0;\ A'x' + b\lambda = b\} = \max\{y^{\mathrm{T}}b \mid y^{\mathrm{T}}A' \leqslant 0,\ y^{\mathrm{T}}b \leqslant 1\}\ . \tag{3.9}$$

If the optimum value is 0, let x_0', λ be an optimum solution for the minimum. So $x_0' \geq 0$, $A'x_0' = b$, and $\lambda = 0$. Hence by adding zero-components, we obtain a vector $x_0 \geq 0$ such that $Ax_0 = b$ and $(x_0)_j = 0$ if $y_0^{\mathrm{T}} a_j < c_j$. By complementary slackness, it follows that x_0 and y_0 are optimum solutions for problems (3.7) and (3.8). If the optimum value in problem (3.9) is positive, it is 1. Let u be an optimum solution for the maximum. Let θ be the largest real number satisfying

$$(y_0 + \theta u)^{\mathrm{T}} A \leq c^{\mathrm{T}} . \tag{3.10}$$

(Note that $\theta > 0$.) Reset $y_0 := y_0 + \theta u$, and start the iteration anew.

This describes the primal–dual method. It reduces problem (3.7) to (3.9), which is often an easier problem, consisting only of testing feasibility of: $x' \geq 0$, $A'x' = b$.

The primal–dual method can equally be considered as a *gradient method*. Suppose we wish to solve problem (3.8), and we have a feasible solution y_0. This y_0 is not optimal if and only if we can find a vector u such that $u^{\mathrm{T}} b > 0$ and u is a *feasible direction* in y_0 [i.e., $(y_0 + \theta u)^{\mathrm{T}} A \leq c^{\mathrm{T}}$ for some $\theta > 0$]. If we let A' consist of those columns of A in which $y_0^{\mathrm{T}} A \leq c^{\mathrm{T}}$ has equality, then u is a feasible direction if and only if $u^{\mathrm{T}} A' \leq 0$. So u can be found by solving the right-hand side of problem (3.9).

Application 3.11 (*Maximum flow*). Let $D = (V, A)$ be a directed graph, let r, $s \in V$, and let a "capacity" function $c: A \to \mathbb{Q}_+$ be given. The *maximum flow problem* is to find the maximum amount of flow from r to s, subject to c:

$$\begin{aligned} \text{maximize} \quad & \sum_{a \in \delta^+(r)} x(a) - \sum_{a \in \delta^-(r)} x(a) \\ \text{subject to} \quad & \sum_{a \in \delta^+(v)} x(a) - \sum_{a \in \delta^-(v)} x(a) = 0 , \quad v \in V,\ v \neq r, s , \\ & 0 \leq x(a) \leq c(a) , \quad a \in A . \end{aligned} \tag{3.12}$$

If we have a feasible solution x_0, we have to find a feasible direction in x_0, i.e., a function $u: A \to \mathbb{R}$ satisfying

$$\begin{aligned} & \sum_{a \in \delta^+(r)} u(a) - \sum_{a \in \delta^-(r)} u(a) > 0 , \\ & \sum_{a \in \delta^+(v)} u(a) - \sum_{a \in \delta^-(v)} u(a) = 0 , \quad v \in V,\ v \neq r, s , \\ & u(a) \geq 0 , \quad a \in A,\ x_0(a) = 0 , \\ & u(a) \leq 0 , \quad a \in A,\ x_0(a) = c(a) . \end{aligned} \tag{3.13}$$

One easily checks that this problem is equivalent to the problem of finding an undirected path from r to s in $D = (V, A)$ so that for any arc a in the path,

$$\begin{aligned}&\text{if } x_0(a) = 0, \quad \text{then arc } a \text{ is traversed forward,}\\ &\text{if } x_0(a) = c(a), \text{ then arc } a \text{ is traversed backward,}\\ &\text{if } 0 < x_0(a) < c(a), \text{ then arc } a \text{ is traversed forward or backward.}\end{aligned} \tag{3.14}$$

If we have found such a path, we find u as in (3.13) (by taking $u(a) = +1$ or -1 if a occurs in the path forward or backward, respectively, and $u(a) = 0$ if a does not occur in the path). Taking the highest θ for which $x_0 + \theta u$ is feasible in problem (3.12) gives us the next feasible solution. The path is called a *flow-augmenting path*, since the new solution has a higher objective value than the old. Iterating this process we finally get an optimum flow. This is exactly Ford and Fulkerson's algorithm (1957) for finding a maximum flow, which is therefore an example of a primal–dual method. [Dinits (1970) and Edmonds and Karp (1972) showed that a version of this algorithm is a polynomial-time method.]

3.3. *The ellipsoid method*

The ellipsoid method, developed by Shor (1970a,b, 1977) and Yudin and Nemirovskiĭ (1976/1977, 1977) for nonlinear programming, was shown by Khachiyan (1979) to solve linear programming in polynomial time. Very roughly speaking, it works as follows.

Suppose we wish to solve the LP problem

$$\max\{c^{\mathrm{T}}x \mid Ax \leqslant b\}\ , \tag{3.15}$$

where $A \in \mathbb{Q}^{m\times n}$, $b \in \mathbb{Q}^m$, and $c \in \mathbb{Q}^n$. Let us assume that the polyhedron $P := \{x \mid Ax \leqslant b\}$ is bounded. Then it is not difficult to calculate a number R such that $P \subseteq \{x \in \mathbb{R}^n \mid \|x\| \leqslant R\}$. We construct a sequence of ellipsoids E_0, E_1, $E_2, \ldots$, each containing the optimum solutions of problem (3.15). First, $E_0 := \{x \in \mathbb{R}^n \mid \|x\| \leqslant R\}$. Suppose ellipsoid E_t has been found. Let z be its center.

If $Az \leqslant b$ does not hold, let $a_k^{\mathrm{T}}x \leqslant b_k$ be an inequality in $Ax \leqslant b$ violated by z. Next let E_{t+1} be the ellipsoid of smallest volume satisfying $E_{t+1} \supseteq E_t \cap \{x \mid a_k^{\mathrm{T}}x \leqslant a_k^{\mathrm{T}}z\}$. If $Az \leqslant b$ does hold, let E_{t+1} be the ellipsoid of smallest volume satisfying $E_{t+1} \supseteq E_t \cap \{x \mid c^{\mathrm{T}}x \geqslant c^{\mathrm{T}}z\}$.

One can prove that these ellipsoids of smallest volume are unique, and that the parameters determining E_{t+1} can be expressed straightforwardly in those determining E_t and in a_k, respectively c. Moreover, $\mathrm{vol}(E_{t+1}) < \mathrm{e}^{-1/3n} \cdot \mathrm{vol}(E_t)$. Hence the volumes of the successive ellipsoids decrease exponentially fast. Since the optimum solutions of problem (3.15) belong to each E_t, we may hope that the centers of the ellipsoids converge to an optimum solution of problem (3.15).

To make this description more precise, an important problem to be solved is that ellipsoids with very small volume can still have a large diameter [so that the centers of the ellipsoids can remain far from any optimum solution of problem (3.15)]. Another, technical, problem is that the unique smallest elipsoid is usually determined by irrational parameters, so that if we work in rational arithmetic we must allow approximations of the successive ellipsoids. These problems can be overcome, and a polynomially bounded running time can be proved.

It was observed by Grötschel et al. (1981), Karp and Papadimitriou (1982) and Padberg and Rao (1980) that in applying the ellipsoid method, it is not necessary that the system $Ax \leq b$ be explicitly given. It suffices to have a "subroutine" to decide whether or not a given vector z belongs to the feasible region of problem (3.15), and to find a separating hyperplane in case z is not feasible. This is especially useful for linear programs coming from combinatorial optimization problems, where the number of inequalities can be exponentially large (in the size of the underlying data-structure), but can yet be tested in polynomial time.

This leads to the following result (Grötschel et al. 1981). Suppose we are given, for each graph $G = (V, E)$, a collection $\mathcal{F}_G$ of subsets of E. For example,

(i) $\mathcal{F}_G$ is the collection of matchings in G;
(ii) $\mathcal{F}_G$ is the collection of spanning trees in G; (3.16)
(iii) $\mathcal{F}_G$ is the collection of Hamiltonian circuits in G.

With any class $(\mathcal{F}_G \mid G$ graph), we can associate the following problem.

Optimization Problem 3.17. Given a graph $G = (V, E)$ and $c \in \mathbb{Q}^E$, find $F \in \mathcal{F}_G$ maximizing $\sum_{e \in F} c_e$.

So if $(\mathcal{F}_G \mid G$ graph) is as in (i), (ii), and (iii) above, Problem 3.17 amounts to the problems of finding a maximum weighted matching, a maximum weighted spanning tree, and a maximum weighted Hamiltonian circuit (the traveling salesman problem), respectively.

The optimization problem is called *solvable in polynomial time*, or *polynomially solvable*, if it is solvable by an algorithm whose running time is bounded by a polynomial in the input size of Problem 3.17, which is $|V| + |E| +$ size(c). Here size(c) $:= \sum_{e \in E}$ size(c_e), where the size of a rational number p/q is $\log_2((|p| + 1) + \log_2(|q|)$. So size($c$) is about the space needed to specify c in binary notation.

Define also the following problem for any fixed class $(\mathcal{F}_G \mid G$ graph).

Separation Problem 3.18. Given a graph $G = (V, E)$ and $x \in \mathbb{Q}^E$, determine whether or not x belongs to conv$\{\chi^F \mid F \in \mathcal{F}_G\}$, and if not, find a separating hyperplane.

Theorem 3.19. *For any fixed class* $(\mathcal{F}_G \mid G$ *graph), the Optimization Problem* 3.17 *is polynomially solvable if and only if the Separation Problem* 3.18 *is polynomially solvable.*

The theorem implies that with respect to the question of polynomial-time solvability, the polyhedral combinatorics approach described in section 1 (i.e., studying the convex hull) is, implicitly or explicitly, unavoidable: a combinatorial optimization problem is polynomially solvable if and only if the corresponding convex hulls can be described decently, in the sense of the polynomial-time solvability of the separation problem. This can be stated also in the negative: if a combinatorial optimization problem is not polynomially solvable (perhaps the traveling salesman problem), then the corresponding polytopes have no such decent description.

The ellipsoid method does not give a practical method, so Theorem 3.19 is more of theoretical value. In some cases, with Theorem 3.19 the polynomial solvability of a combinatorial optimization problem was proved, and that then formed a motivation for finding a practical polynomial-time algorithm for the problem.

One drawback of the ellipsoid method is that the number of ellipsoids to be evaluated depends on the size of the objective vector c. This does not conflict with the definition of polynomial solvability, but is not very attractive in practice. It would be preferable for the size of c only to influence the sizes of the numbers occurring when we perform the algorithm, but not the number of arithmetic operations to be performed. An algorithm for Optimization Problem 3.17 is called *strongly polynomial* if it consists of a number of arithmetic operations, bounded by a polynomial in $|V| + |E|$, on numbers of size bounded by a polynomial in $|V| + |E| + \text{size}(c)$. Such an algorithm is obviously polynomial-time.

Interestingly, Frank and Tardos (1985) showed, with the help of the "basis reduction method" (Lenstra et al. 1982):

Theorem 3.20. *For any fixed class* ($\mathcal{F}_G \mid G$ graph), *if there exists a polynomial-time algorithm for Optimization Problem* 3.17, *then there exists a strongly polynomial algorithm for it.*

At the moment of writing, it is not yet clear whether this result leads to practical algorithms.

Finally we note that it is not necessary to restrict $\mathcal{F}_G$ to collections of subsets of the edge set E. For instance, similar results hold if we consider collections $\mathcal{F}_G$ of subsets of the vertex set V. Moreover, we can consider classes $(\mathcal{F}_G \mid G \in \mathcal{G})$, where $\mathcal{G}$ is a subcollection of the set of all graphs. Similarly, we can consider classes ($\mathcal{F}_D \mid D$ directed graph), ($\mathcal{F}_H \mid H$ hypergraph), ($\mathcal{F}_M \mid M$ matroid), and so on.

More on the ellipsoid method can be found in Grötschel et al. (1988).

We finally mention the method of Karmarkar (1984) for linear programming; this appears to be competitive with the simplex method, but its impact on polyhedral combinatorics is not yet clear at the moment of writing.

4. Total unimodularity

A matrix is called *totally unimodular* if each subdeterminant belongs to $\{0, +1, -1\}$. In particular, each entry of a totally unimodular matrix belongs to $\{0, +1,$

$-1\}$. The importance of total unimodularity for polyhedral combinatorics comes from the following theorem (Hoffman and Kruskal 1956).

Theorem 4.1. *Let A be a totally unimodular $m \times n$ matrix and let $b \in \mathbb{Z}^m$. Then the polyhedron $P := \{x \mid Ax \leqslant b\}$ is integral.*

Proof. Let $F = \{x \mid A'x = b'\}$ be a minimal face of P, where $A'x \leqslant b'$ is a subsystem of $Ax \leqslant b$. Without loss of generality, $A' = [A_1 \; A_2]$, with A_1 nonsingular. Then A_1^{-1} is an integral matrix (as $\det A_1 = \pm 1$), and hence the vector

$$x := \begin{pmatrix} A_1^{-1}b' \\ 0 \end{pmatrix} \tag{4.2}$$

is an integral vector in F. □

In fact, Hoffman and Kruskal showed that an integral $m \times n$ matrix A is totally unimodular if and only if for each $b \in \mathbb{Z}^m$, each vertex of the polyhedron $\{x \in \mathbb{R}^n \mid x \geqslant 0, Ax \leqslant b\}$ is integral.

We mention a strengthening of Theorem 4.1 due to Baum and Trotter (1977). A polyhedron P in $\mathbb{R}^n$ is said to have the *integer decomposition property* if for each $k \in \mathbb{N}$ and for each integral vector z in kP $(=\{kx \mid x \in P\})$, there exist integral vectors $x_1, \ldots, x_k$ in P so that $z = x_1 + \cdots + x_k$. It is not difficult to see that each polyhedron with the integer decomposition property is integral.

Theorem 4.3. *Let A be a totally unimodular $m \times n$ matrix and let $b \in \mathbb{Z}^m$. Then the polyhedron $P := \{x \mid Ax \leqslant b\}$ has the integer decomposition property.*

Proof. Let $k \in \mathbb{N}$ and $z \in kP \cap \mathbb{Z}^n$. By induction on k we show that $z = x_1 + \cdots + x_k$ for integral vectors $x_1, \ldots, x_k$ in P. By Theorem 4.1, there exists an integral vector, say x_k, in the polyhedron $\{x \mid Ax \leqslant b, -Ax \leqslant (k-1)b - Az\}$ [since (i) the constraint matrix $\left[\begin{smallmatrix} A \\ -A \end{smallmatrix}\right]$ is totally unimodular, (ii) the right-hand-side vector $\left(\begin{smallmatrix} b \\ (k-1)b - Az \end{smallmatrix}\right)$ is integral, and (iii) the polyhedron is nonempty, as it contains $k^{-1}z$]. Then $z - x_k \in (k-1)P$, whence by induction $z - x_k = x_1 + \cdots + x_{k-1}$ for integral vectors $x_1, \ldots, x_{k-1}$ in P. □

The following theorem collects together several other characterizations of total unimodularity.

Theorem 4.4. *Let A be a matrix with entries* 0, $+1$, *and* -1. *Then the following characterizations are equivalent*:

(i) *A is totally unimodular, i.e., each square submatrix of A has determinant in* $\{0, +1, -1\}$;

(ii) *each collection of columns of A can be split into two parts so that the sum of the columns in one part, minus the sum of the columns in the other part, is a vector with entries* 0, $+1$, *and* -1 *only*;

(iii) *each nonsingular submatrix of A has a row with an odd number of nonzero components*;

(iv) *the sum of the entries in any square submatrix of A with even row and column sums, is divisible by four*;

(v) *no square submatrix of A has determinant $+2$ or -2.*

Characterization (ii) is due to Ghouila-Houri (1962), (iii) and (iv) to Camion (1965), and (v) to R. E. Gomory (cf. Camion 1965).

There are several further characterizations of total unimodularity. By far the deepest is due to Seymour (1980) (see chapter 10). For an efficient algorithm to test total unimodularity, see Truemper (1982). See also Truemper (1990).

4.1. Application: bipartite graphs

It is not difficult to see that the $V \times E$ incidence matrix A of a bipartite graph $G = (V, E)$ is totally unimodular: any square submatrix B of A either has a column with at most one 1 (in which case $\det B \in \{0, \pm 1\}$ by induction), or has two 1's in each column (in which case $\det B = 0$ by the bipartiteness of G). In fact, the incidence matrix of a graph G is totally unimodular if and only if G is bipartite.

The total unimodularity of the incidence matrix of a bipartite graph has several consequences, some of which we will describe now.

Definition 4.5. The *matching polytope* of a graph $G = (V, E)$ is the polytope $\operatorname{conv}\{\chi^M \mid M \text{ matching}\}$ in $\mathbb{R}^E$. Theorem 4.1 directly implies that the matching polytope of a bipartite graph G is equal to the set of all vectors x in $\mathbb{R}^E$ satisfying

$$\begin{aligned} &\text{(i)} \quad x_e \geq 0\,, \quad e \in E\,, \\ &\text{(ii)} \quad \sum_{e \ni v} x_e \leq 1\,, \quad v \in V \end{aligned} \tag{4.6}$$

[since the polyhedron determined by (4.6) is integral].

Clearly, the matching polytope of $G = (V, E)$ has dimension $|E|$. Each inequality in (4.6) is facet-determining, except if G has a vertex of degree at most 1. It is not difficult to see that the incidence vectors χ^M, $\chi^{M'}$ of two matchings M, M' are adjacent on the matching polytope iff $M \,\Delta\, M'$ is a path or circuit, where Δ denotes symmetric difference. Hence, the matching polytope of G has diameter at most $\nu(G)$. (This paragraph holds also for nonbipartite graphs.)

The above characterization of the matching polytope for bipartite graphs implies that for any bipartite graph $G = (V, E)$ and any "weight" function $c: E \rightarrow \mathbb{R}_+$:

$$\text{maximum weight of a matching} = \max\{c^{\mathrm{T}}x \mid x \geq 0, Ax \leq \mathbf{1}\}\,, \tag{4.7}$$

where A is the incidence matrix of A, $\mathbf{1}$ denotes an all-one column vector, and where the *weight* of a set is the sum of the weights of its elements. In particular,

$$\nu(G) = \max\{\mathbf{1}^{\mathrm{T}}x \mid x \geq 0, Ax \leq \mathbf{1}\}\ . \tag{4.8}$$

Definition 4.9. The *node-cover polytope* of a graph $G = (V, E)$ is the polytope $\mathrm{conv}\{\chi^N \mid N$ node cover$\}$ in $\mathbb{R}^V$. Again, Theorem 4.1 implies that, if G is bipartite, the node-cover polytope of G is equal to the set of all vectors y in $\mathbb{R}^V$ satisfying

$$\begin{aligned} &\text{(i)} \quad 0 \leq y_v \leq 1\ , \quad v \in V\ , \\ &\text{(ii)} \quad y_v + y_w \geq 1\ , \quad \{v, w\} \in E\ . \end{aligned} \tag{4.10}$$

It follows that for any weight function $w: V \rightarrow \mathbb{R}_+$:

$$\text{minimum weight of a node cover} = \min\{w^{\mathrm{T}}y \mid y \geq 0, y^{\mathrm{T}}A \geq \mathbf{1}\}\ , \tag{4.11}$$

where A again is the $V \times E$ incidence matrix of G. In particular,

$$\tau(G) = \min\{\mathbf{1}^{\mathrm{T}}y \mid y \geq 0, y^{\mathrm{T}}A \geq \mathbf{1}\}\ . \tag{4.12}$$

Now, by linear programming duality, we know that problems (4.8) and (4.12) are equal, i.e., we have *König's Matching Theorem*: $\nu(G) = \tau(G)$ for bipartite G.

By Theorem 4.3, the matching polytope P of G has the integer decomposition property. This has the following consequence. Let $k := \Delta(G)$ (the maximum degree of G). Then $(1, \ldots, 1)^{\mathrm{T}} \in \mathbb{R}^E$ belongs to kP, and hence is the sum of k integer vectors in P. Each of these vectors being the incidence vector of a matching, it follows that E can be partitioned into k matchings. So we have *König's Edge-Coloring Theorem*: the edge-coloring number $\gamma(G)$ of a bipartite graph G is equal to its maximum degree.

We briefly mention some more examples of the consequences of Theorems 4.1 and 4.3 to bipartite graphs.

Definition 4.13. The *perfect matching polytope* of a graph $G = (V, E)$ is the polytope $\mathrm{conv}\{\chi^M \mid M$ perfect matching$\}$ in $\mathbb{R}^E$. It is a face of the matching polytope of G. For bipartite graphs, by (4.6), the perfect matching polytope is determined by

$$\begin{aligned} &\text{(i)} \quad x_e \geq 1\ , \quad e \in E\ , \\ &\text{(ii)} \quad \sum_{e \ni v} x_e = 1\ , \quad v \in V\ . \end{aligned} \tag{4.14}$$

This is equivalent to a theorem of Birkhoff (1946): each doubly stochastic matrix is a convex combination of permutation matrices.

One easily checks that the incidence vectors χ^M, $\chi^{M'}$ of two perfect matchings M, M' are adjacent on the perfect matching polytope if and only if $M \triangle M'$ is a circuit (cf. Balinski and Russakoff 1974). So the perfect matching polytope has diameter at most $\frac{1}{2}|\mathrm{V}|$. The dimension of the perfect matching polytope of a bipartite graph is equal to $|E'| - |V| + 1$, where $E' := \bigcup M \setminus (\bigcap M)$, where the union and intersection both range over all perfect matchings (see Lovász and Plummer 1986).

Definition 4.15. The *assignment polytope* of order n is the perfect matching polytope of $K_{n,n}$. Equivalently, it is the polytope in $\mathbb{R}^{n \times n}$ of all matrices $(x_{ij})_{i,j=1}^{n}$ satisfying

$$\begin{aligned} &\text{(i)} \quad x_{ij} \geq 0\,, \quad i, j = 1, \ldots, n\,, \\ &\text{(ii)} \quad \sum_{i=1}^{n} x_{ij} = 1\,, \quad j = 1, \ldots, n\,, \\ &\text{(iii)} \quad \sum_{j=1}^{n} x_{ij} = 1\,, \quad i = 1, \ldots, n\,. \end{aligned} \tag{4.16}$$

(Such matrices are called *doubly stochastic*.)

Balinski and Russakoff (1974) studied assignment polytopes, proving inter alia that they have diameter 2 (if $n \geq 4$). See also Balinski (1985), Bertsekas (1981), Goldfarb (1985), Hung (1983), Padberg and Rao (1974), and Roohy-Laleh (1981).

Definition 4.17. The *stable-set polytope* of a graph $G = (V, E)$ is the polytope $\operatorname{conv}\{\chi^C \mid C \text{ stable set}\}$ in $\mathbb{R}^V$. By Theorem 4.1, for bipartite G, it is determined by

$$\begin{aligned} &\text{(i)} \quad 0 \leq y_v \leq 1\,, \quad v \in V\,, \\ &\text{(ii)} \quad y_v + y_w \leq 1\,, \quad \{v, w\} \in E\,. \end{aligned} \tag{4.18}$$

So if A is the $V \times E$ incidence matrix of the bipartite graph G, and $w: V \rightarrow \mathbb{R}_+$ is a "weight" function, then

$$\text{maximum weight of a stable set} = \max\{w^{\mathrm{T}}y \mid y \geq 0,\ y^{\mathrm{T}}A \leq \mathbf{1}^{\mathrm{T}}\}\,. \tag{4.19}$$

In particular:

$$\alpha(G) = \max\{\mathbf{1}^{\mathrm{T}}y \mid y \geq 0,\ y^{\mathrm{T}}A \leq \mathbf{1}^{\mathrm{T}}\}\,. \tag{4.20}$$

Definition 4.21. The *edge-cover polytope* of a graph $G = (V, E)$ is the polytope $\operatorname{conv}\{\chi^F \mid F \text{ edge cover}\}$ in $\mathbb{R}^E$. By Theorem 4.1, for bipartite G, it is determined by

$$\begin{aligned} &\text{(i)} \quad 0 \leq x_e \leq 1\,, \quad e \in E\,, \\ &\text{(ii)} \quad \sum_{e \ni v} x_e \geq 1\,, \quad v \in V\,, \end{aligned} \tag{4.22}$$

assuming G has no isolated vertices. Hurkens (1991) characterized adjacency on the edge-cover polytope, and showed that its diameter is $|E| - \rho(G)$.

From (4.22) it follows that for any "weight" function $w: E \to \mathbb{R}_+$,

$$\text{minimum weight of an edge cover} = \min\{w^{\mathrm{T}}x \mid x \geq 0,\ Ax \geq \mathbf{1}\}\,. \tag{4.23}$$

In particular:

$$\rho(G) = \min\{\mathbf{1}^{\mathrm{T}}x \mid x \geq 0,\ Ax \geq \mathbf{1}\}\,. \tag{4.24}$$

By linear programming duality, (4.20) and (4.24) are equal, and hence we have *König's Covering Theorem*: $\alpha(G) = \rho(G)$ for bipartite G.

By Theorem 4.3, the edge-cover polytope of a bipartite graph has the integer decomposition property, implying a result of Gupta (1967): the maximum number of pairwise disjoint edge covers in a bipartite graph is equal to its minimum degree.

Let A be the incidence matrix of the bipartite graph $G = (V, E)$, let $w \in \mathbb{Z}^E$, $b \in \mathbb{Z}^V$, and consider the linear programs in the following duality equations:

$$\begin{array}{ll} \text{(i)} & \max\{w^{\mathrm{T}}x \mid x \geq 0,\ Ax \leq b\} = \min\{y^{\mathrm{T}}b \mid y \geq 0,\ y^{\mathrm{T}}A \geq w^{\mathrm{T}}\}\,, \\ \text{(ii)} & \min\{w^{\mathrm{T}}x \mid x \geq 0,\ Ax \geq b\} = \max\{y^{\mathrm{T}}b \mid y \geq 0,\ y^{\mathrm{T}}A \leq w^{\mathrm{T}}\}\,. \end{array} \tag{4.25}$$

By Theorem 4.1, these programs have integer optimum solutions. The special case $b = \mathbf{1}$ is equivalent to the following min–max relations of Egerváry (1931):

(i) the maximum weight of a matching is equal to the minimum value of $\sum_{v \in V} y_v$, where $y: V \to \mathbb{Z}_+$ such that $y_u + y_v \geq w_e$ $\forall e = \{u, v\} \in E$;

(4.26)

(ii) The minimum weight of an edge cover is equal to the maximum value of $\sum_{v \in V} y_v$, where $y: V \to \mathbb{Z}_+$ such that $y_u + y_v \leq w_e$ $\forall e = \{u, v\} \in E$.

Definition 4.27. The *transportation polytope* for $a \in \mathbb{R}^m_+$, $b \in \mathbb{R}^n_+$ is the set of all vectors $(x_{ij} \mid i = 1, \ldots, m,\ j = 1, \ldots, n)$ in $\mathbb{R}^{m \times n}$ satisfying

$$\begin{array}{ll} \text{(i)} & x_{ij} \geq 0\,, \quad i = 1, \ldots, m,\ j = 1, \ldots, n\,, \\ \text{(ii)} & \sum_{j=1}^{n} x_{ij} = a_i\,, \quad i = 1, \ldots, m\,, \\ \text{(iii)} & \sum_{i=1}^{n} x_{ij} = b_j\,, \quad j = 1, \ldots, n\,. \end{array} \tag{4.28}$$

It is related to the *Hitchcock–Koopmans transportation problem* (Hitchcock 1941, Koopmans 1948). Klee and Witzgall (1968) studied transportation polytopes, showing that x satisfying (4.28) is a vertex iff $\{\{p_i, q_j\} \mid x_{ij} > 0\}$ contains no circuits (where $p_1, \ldots, p_m, q_1, \ldots, q_n$ are vertices). Moreover, the dimension is $(m-1)\ (n-1)$ if a and b are positive (if the polytope is nonempty, i.e., if $\sum_i a_i = \sum_j b_j$). Bolker (1972) and Balinski (1974) showed the Hirsch Conjecture for some classes of transportation polytopes. Bolker (1972) and Ahrens (1981) studied the number of vertices of transportation polytopes.

Definition 4.29. Related is the *dual transportation polyhedron*, which is, for fixed $c \in \mathbb{R}^{m \times n}$, defined as the set of all vectors $(u; v)$ in $\mathbb{R}^m \times \mathbb{R}^n$ satisfying

$$u_1 = 0 , \qquad u_i + v_j \geq c_{ij} , \quad i = 1, \ldots , m, \ j = 1, \ldots , n . \tag{4.30}$$

It is not difficult to see that the dimension is $m + n - 1$, and that $(u; v)$ satisfying (4.30) is a vertex iff $\{\{p_i, q_j\} \mid u_i + v_j = c_{ij}\}$ is a connected graph on vertex set $\{p_1, \ldots , p_m, q_1, \ldots , q_n\}$. Balinski (1984) showed that the diameter of (4.30) is at most $(m - 1)$ $(n - 1)$, thus proving the Hirsch Conjecture for this class of polyhedra. Balinski and Russakoff (1984) made a further study of dual transportation polyhedra, characterizing vertices and higher-dimensional faces by means of partitions. See also Balinski (1983), Ikura and Nemhauser (1983), and Zhu (1963).

4.2. Application: directed graphs

Total unimodularity also implies several results for flows and circulations in directed graphs. Let M be the $V \times A$ incidence matrix of a digraph $D = (V, A)$. Then M is totally unimodular. Again this can be shown by induction: let B be a square submatrix of M. If B has a column with at most one nonzero, then $\det B \in \{0, \pm 1\}$ by induction. If each column of B contains a $+1$ and a -1, then $\det B = 0$.

There are the following consequences.

Definition 4.31. Let $D = (V, A)$ be a digraph, let $r, s \in V$, and let $c \in \mathbb{R}^A_+$ be a "capacity" function. Then the *r–s-flow polytope* is the set of all vectors x in $\mathbb{R}^A$ satisfying

$$\begin{aligned} &\text{(i)} \quad 0 \leq x_a \leq c_a , \quad a \in A , \\ &\text{(ii)} \quad \sum_{a \in \delta^-(v)} x_a = \sum_{a \in \delta^+(v)} x_a , \quad v \in V, \ v \neq r, s . \end{aligned} \tag{4.32}$$

Any vector x satisfying (4.32) is called an *r–s-flow* (*under* c). By the total unimodularity of the incidence matrix of D, if c is integral, then the r–s-flow polytope has integral vertices. Hence, if c is integral, the maximum *value* $(:= \sum_{a \in \delta^+(r)} x_a - \sum_{a \in \delta^-(r)} x_a)$ of an r–s-flow under c is attained by an integral vector (Dantzig 1951b).

4.33 (*Max-Flow Min-Cut Theorem*). By LP duality, the maximum value of an r–s-flow under c is equal to the minimum value of $\sum_{a \in A} y_a c_a$, where $y \in \mathbb{R}^A_+$ is such that there exists a vector z in $\mathbb{R}^V$ satisfying

$$\begin{aligned} &\text{(i)} \quad y_a - z_v + z_w \geq 0 , \quad a = (v, w) \in A , \\ &\text{(ii)} \quad z_r = 1 , \qquad z_s = 0 . \end{aligned} \tag{4.34}$$

Again, by the total unimodularity of the incidence matrix of D, we may take the minimizing y, z to be integral. Let $W := \{v \in V \mid z_v \geq 1\}$. Then for $a = (v, w) \in$

$\delta^+(W)$ we have $y_a \geqslant z_v - z_w \geqslant 1$, and hence

$$\sum_{a\in A} y_a c_a \geqslant \sum_{a\in\delta\ \ (W)} y_a c_a \geqslant \sum_{a\in\delta\ \ (W)} c_a . \tag{4.35}$$

So the maximum flow value is not less than the capacity of cut $\delta^+(W)$. Since it also cannot be larger, we have Ford and Fulkerson's Max-Flow Min-Cut Theorem.

Definition 4.36. Given digraph $D = (V, A)$ and $r, s \in V$, the *shortest-path polytope* is the convex hull of all incidence vectors χ^P of subsets P of A, being a disjoint union of an r–s-path and some directed circuits. By the total unimodularity of the incidence matrix of D, this polytope is equal to the set of all vectors $x \in \mathbb{R}^A$ satisfying

$$\begin{aligned}
&\text{(i)} && 0 \leqslant x_a \leqslant 1 , \quad a \in A ,\\
&\text{(ii)} && \sum_{a\in\delta^+(v)} x_a = \sum_{a\in\delta^-(v)} x_a , \quad v \in V,\ v \neq r, s ,\\
&\text{(iii)} && \sum_{a\in\delta^+(r)} x_a - \sum_{a\in\delta^-(r)} x_a = 1 .
\end{aligned} \tag{4.37}$$

So it is the intersection of an r–s-flow polytope with the hyperplane determined by (iii). Saigal (1969) showed that the Hirsch Conjecture holds for the class of shortest-path polytopes.

Definition 4.38. For digraph $D = (V, A)$ and $l, u \in \mathbb{R}^A$, the *circulation polytope* is the set of all *circulations* between l and u, i.e., vectors $x \in \mathbb{R}^A$ satisfying

$$\begin{aligned}
&\text{(i)} && l_a \leqslant x_a \leqslant u_a , \quad a \in A ,\\
&\text{(ii)} && Mx = 0 ,
\end{aligned} \tag{4.39}$$

where M is the incidence matrix of D. By the total unimodularity of M, if l and u are integral, then the circulation polytope is integral. So if l and u are integral, and there exists a circulation, there exists an integral circulation. Similarly, a minimum-cost circulation can be taken to be integral.

By Farkas's Lemma, the circulation polytope is nonempty iff there are no vectors z, $w \in \mathbb{R}^A$, $y \in \mathbb{R}^V$ satisfying

$$\begin{aligned}
&\text{(i)} && z, w \geqslant 0 ,\\
&\text{(ii)} && z - w + M^{\mathrm{T}} y = 0 ,\\
&\text{(iii)} && u^{\mathrm{T}} z - l^{\mathrm{T}} w < 0 .
\end{aligned} \tag{4.40}$$

Suppose now $l \leqslant u$, and (4.40) has a solution. Then there is also a solution satisfying $\mathbf{0} \leqslant y \leqslant \mathbf{1}$, and hence, by the total unimodularity of M, there is a solution z, w, y with y a $\{0, 1\}$-vector. We may assume that $z_a w_a = 0$ for each arc

a. Then, for $W := \{v \in V \mid y_v = 1\}$,

$$\sum_{a\in\delta^-(W)} u_a - \sum_{a\in\delta^+(W)} l_a = u^{\mathrm{T}}z - l^{\mathrm{T}}w < 0 . \tag{4.41}$$

Thus we have Hoffman's *Circulation Theorem* (Hoffman 1960): there exists a circulation x satisfying $l \leqslant x \leqslant u$ iff $l \leqslant u$ and there is no subset W of V with $\sum_{a\in\delta^-(W)} u_a < \sum_{a\in\delta^+(W)} l_a$.

4.42. More generally, for $l, u \in \mathbb{R}^A$ and $b', b'' \in \mathbb{R}^V$, the polyhedron

$$\{x \in \mathbb{R}^A \mid l \leqslant x \leqslant u, b' \leqslant Mx \leqslant b''\} \tag{4.43}$$

is integral, if l, u, b' and b'' are integral. Moreover, the total unimodularity of M yields a characterization of the nonemptiness of the polyhedron (4.43), extending Hoffman's Circulation Theorem.

It is not difficult to see that (4.43) is an affine transformation of the polytope of vectors $(x'; x''; y'; y'')$ in $\mathbb{R}^A \times \mathbb{R}^A \times \mathbb{R}^V \times \mathbb{R}^V$ satisfying

$$\begin{aligned}
&x'_a \geqslant 0, \quad x''_a \geqslant 0, \qquad a \in A,\\
&y'_v \geqslant 0, \quad y''_v \geqslant 0, \qquad v \in V,\\
&\sum_{a\in\delta^+(v)} x'_a + \sum_{a\in\delta^-(v)} x''_a + y'_v = b''_v + \sum_{a\in\delta^-(v)} u_a - \sum_{a\in\delta^+(v)} l_a, \quad v \in V,\\
&x'_a + x''_a = u_a - l_a, \quad a \in A,\\
&y'_v + y''_v = b''_v - b'_v, \quad v \in V
\end{aligned} \tag{4.44}$$

(the transformation is given by $x_a := x'_a + l_a$). Thus (4.43) is transformed into a face of the transportation polytope (4.27). In this way, several results for (4.43) can be derived from results for transportation polytopes.

Let $D = (V, A)$ be a directed graph, and let $T \subseteq A$ be a spanning tree in D. Consider the $T \times (A\backslash T)$ matrix N defined, for $a \in T$ and $a' = (v, w) \in A\backslash T$, by:

$$N_{a',a} := \begin{cases} 0 & \text{if } a \text{ does not occur in the } v\text{–}w \text{ path in } T,\\ -1 & \text{if } a \text{ occurs forward in the } v\text{–}w \text{ path in } T,\\ +1 & \text{if } a \text{ occurs backward in the } v\text{–}w \text{ path in } T. \end{cases} \tag{4.45}$$

Then N is totally unimodular, as can be seen with the help of Ghouila-Houri's characterization (4.4) (ii). A vector $x = \binom{x'}{x''}$ in $\mathbb{R}^{A\backslash T} \times \mathbb{R}^T$ satisfies $Mx = 0$ (where M is the incidence matrix of D) if and only if $x'' = Nx'$. Thus (4.39) can be reformulated as

$$\begin{aligned}
&l_a \leqslant x'_a \leqslant u_a, \quad a \in A\backslash T,\\
&l_a \leqslant (Nx')_a \leqslant u_a, \quad a \in T.
\end{aligned} \tag{4.46}$$

By the total unimodularity of N, the polytope determined by (4.46) has integer vertices, if all l_a and u_a are integer.

A special case is formed by the $\{0, 1\}$-matrices with the *consecutive ones property*: in each column, the 1's form an interval (fixing some ordering of the rows, as usual). This special case arises when T is a directed path, and each arc in $A\backslash T$ forms a directed circuit with some subpath in T.

For related results, see also Hoffman (1960, 1979).

5. Total dual integrality

Total dual integrality appears to be a powerful technique in deriving min–max relations and the integrality of polyhedra. It is based on the following result, shown, implicitly or explicitly, by Gomory (1963), Lehman (1965), Fulkerson (1971), Chvátal (1973a), Hoffman (1974) and Lovász (1976) for pointed polyhedra, and by Edmonds and Giles (1977) for general polyhedra.

Theorem 5.1. *A rational polyhedron P is integral if and only if each rational supporting hyperplane of P contains integral vectors.*

Proof. Since the intersection of a supporting hyperplane with P is a face of P, necessity of the condition is trivial. To prove sufficiency, suppose that each rational supporting hyperplane of P contains integral vectors. Let $P = \{x \mid Ax \leqslant b\}$, with A and b integral. Let $F = \{x \mid A'x = b'\}$ be a minimal face of P, where $A'x \leqslant b'$ is a subsystem of $Ax \leqslant b$. If F does not contain any integral vector, there exists a vector y such that $c^{\mathrm{T}} := y^{\mathrm{T}}A'$ is an integral vector, while $\delta := y^{\mathrm{T}}b'$ is not an integer (this follows, e.g., from Hermite's Normal Form Theorem). We may suppose that all entries in y are nonnegative (we may replace each entry y_i of y by $y_i - \lfloor y_i \rfloor$). Now $H := \{x \mid c^{\mathrm{T}}x = \delta\}$ is a supporting hyperplane of P, not containing any integral vector. □

Note that the special case where P is pointed can be shown without appealing to Hermite's Theorem: if x^* is a nonintegral vertex of P, w.l.o.g. $x^*_1 \not\in \mathbb{Z}$. There exist supporting hyperplanes $H = \{x \mid c^{\mathrm{T}}x = c^{\mathrm{T}}x^*\}$ and $\tilde{H} = \{x \mid \tilde{c}^{\mathrm{T}}x = \tilde{c}^{\mathrm{T}}x^*\}$ touching P in x^* such that c and $\tilde{c}$ are integral and such that $c^{\mathrm{T}} - \tilde{c}^{\mathrm{T}} = (1, 0, \ldots, 0)$. If both H and $\tilde{H}$ contain integral vectors, we know $c^{\mathrm{T}}x^* \in \mathbb{Z}$ and $\tilde{c}^{\mathrm{T}}x^* \in \mathbb{Z}$. However, $(c - \tilde{c})^{\mathrm{T}}x^* = x^*_1 \not\in \mathbb{Z}$.

Theorem (5.1) can be applied as follows. Consider the LP problem

$$\max\{c^{\mathrm{T}}x \mid Ax \leqslant b\}\,, \tag{5.2}$$

for rational matrix A and rational vectors b, c.

Corollary 5.3. *The following are equivalent*:

(i) *the maximum value in* (5.2) *is an integer for each integral vector c for which the maximum is finite*;

(ii) *the maximum* (5.2) *is attained by an integral optimum solution for each rational vector c for which the maximum is finite*;

(iii) *the polyhedron* $\{x \mid Ax \leqslant b\}$ *is integral.*

Now consider the LP-duality equation

$$\max\{c^{\mathrm{T}}x \mid Ax \leqslant b\} = \min\{y^{\mathrm{T}}b \mid y \geqslant 0, y^{\mathrm{T}}A = c^{\mathrm{T}}\}\,. \tag{5.4}$$

Clearly, we may derive that the maximum value is an integer if we know that the minimum has an integral optimum solution and b is integral. This motivated Edmonds and Giles (1977) to define a system $Ax \leqslant b$ of linear inequalities to be *totally dual integral* (*TDI*) if for each integral vector c, the minimum in (5.4) is attained by an integral optimum solution. Then we have the following consequence.

Corollary 5.5. *Let* $Ax \leqslant b$ *by a system of linear inequalities, with* A *rational and* b *integral. If* $Ax \leqslant b$ *is TDI* (*i.e., the minimum in* (5.4) *is attained by an integral optimum solution* y, *for each integral vector* c *for which the minimum is finite*), *then* $\{x \mid Ax \leqslant b\}$ *is integral* (*i.e., the maximum in* (5.4) *is attained by an integral optimum solution* x, *for each* c *for which the maximum is finite*).

Note that the notion of total dual integrality is *not* symmetric in objective function c and right-hand-side vector b. Indeed, the implication in Corollary 5.5 cannot be reversed: the system $x_1 \geqslant 0$, $x_1 + 2x_2 \geqslant 0$ determines an integral polyhedron in $\mathbb{R}^2$, while it is not TDI. However, Giles and Pulleyblank (1979) showed that if P is an integral polyhedron, then $P = \{x \mid Ax \leqslant b\}$ for some TDI-system $Ax \leqslant b$ with b integral. In Schrijver (1981) it is shown that if P is moreover full-dimensional, then there is a unique minimal TDI-system determining P with A and b integral (minimal under deleting inequalities).

Related to total dual integrality is the notion of *Hilbert basis*: This is a collection $\{a_1, \ldots, a_m\}$ of vectors with the property that if an integer vector x is a nonnegative linear combination of the vectors $a_1, \ldots, a_m$, then it is an *integer* nonnegative linear combination of them.

The relation to total dual integrality is as follows. Let $Ax \leqslant b$ be a system of linear inequalities, and set $P := \{x \mid Ax \leqslant b\}$. If $a^{\mathrm{T}}x \leqslant \beta$ is an inequality from $Ax \leqslant b$ and F is a face of P, we say a is *tight* in F if $a^{\mathrm{T}}x = \beta$ for all x in F. Now $Ax \leqslant b$ is TDI if and only if for each face F of P, the rows of A that are tight in A form a Hilbert basis.

It was shown by Cook et al. (1986a) that if $\{a_1, \ldots, a_m\}$ is a Hilbert basis consisting of integer vectors in $\mathbb{R}^n$, then any integer vector x that is a nonnegative linear combination of $a_1, \ldots, a_m$ is in fact an integer nonnegative linear combination of at most $2n - 1$ of these vectors.

As a consequence one has that if $Ax \leqslant b$ is TDI (in n variables, say), and A is integral, then for any $c \in \mathbb{Z}^n$, $\min\{y^{\mathrm{T}}b \mid y \geqslant 0, y^{\mathrm{T}}A = c^{\mathrm{T}}\}$ is attained by an integer vector y with at most $2n - 1$ nonzero components (if the minimum is finite).

For more on total dual integrality, see Cook (1983, 1986), Edmonds and Giles (1984), and Cook et al. (1984).

We now consider some combinatorial applications of total dual integrality.

Application 5.6 (*Arborescences*). Let $D=(V,A)$ be a directed graph, and let r be a fixed vertex of D. An *r-arborescence* is a set A' of $|V|-1$ arcs forming a spanning tree such that each vertex $v \neq r$ is entered by exactly one arc in A'. So for each vertex v there is a unique directed path in A' from r to v. An *r-cut* is an arc set of the form $\delta^-(U)$, for some nonempty subset U of $V \backslash \{r\}$. As usual $\delta^-(U)$ denotes the set of arcs entering U.

It is not difficult to see that r-arborescences are the inclusion-wise minimal sets of arcs intersecting r-cuts. Conversely, the inclusion-wise minimal r-cuts are the inclusion-wise minimal sets of arcs intersecting all r-arborescences.

Fulkerson (1974) showed:

Fulkerson's Optimum Arborescence Theorem 5.7. *For any "length" function $l: A \rightarrow \mathbb{Z}_+$, the minimum length of an r-arborescence is equal to the maximum number t of r-cuts $C_1, \ldots, C_t$ (repetition allowed) so that no arc a is in more than $l(a)$ of these cuts.*

This result can be formulated in polyhedral terms as follows. Let C be the matrix whose rows are the incidence vectors of all r-cuts. So the columns of C are indexed by A, and the rows by the collection $\mathcal{H} := \{U \mid \emptyset \neq U \subseteq V \backslash \{r\}\}$. Then Theorem 5.7 is equivalent to both optima in the LP-duality equation

$$\min\{l^{\mathrm{T}}x \mid x \geq 0, Cx \geq \mathbf{1}\} = \max\{y^{\mathrm{T}}\mathbf{1} \mid y \geq 0, y^{\mathrm{T}}C \leq l^{\mathrm{T}}\} \tag{5.8}$$

having integral optimum solutions, for each $l \in \mathbb{Z}_+^A$. So in order to show the theorem, by (5.5) it suffices to show that the maximum in (5.8) has an integral optimum solution, for each $l: A \rightarrow \mathbb{Z}$, i.e., that the system $x \geq 0$, $Cx \geq \mathbf{1}$ is TDI. This can be proved as follows (Edmonds and Giles 1977).

Proof of Theorem 5.7. Note that the matrix C is generally not totally unimodular. However, in order to prove that the maximum (5.8) has an integer optimum solution, it suffices to show that there exists a "basis" that is totally unimodular and that attains the maximum. That is, it is enough to find a totally unimodular submatrix C' of C (consisting of rows of C) such that

$$\max\{y^{\mathrm{T}}\mathbf{1} \mid y \geq 0, y^{\mathrm{T}}C \leq l^{\mathrm{T}}\} = \max\{z^{\mathrm{T}}\mathbf{1} \mid z \geq 0, z^{\mathrm{T}}C' \leq l^{\mathrm{T}}\}\,. \tag{5.9}$$

Since the second maximum is attained by an integer optimum solution z (by the total unimodularity of C'), extending z by 0's in the appropriate positions gives an integer optimum solution y for the first maximum.

How can we find such a C'? The key observation is the following. Call a subcollection $\mathcal{F}$ of $\mathcal{H}$ *laminar* if for all T, $U \in \mathcal{F}$ one has $T \subseteq U$ or $U \subseteq T$ or $T \cap U = \emptyset$. Then, if C' is the matrix consisting of the rows of C with index in some laminar family $\mathcal{F}$, C' is totally unimodular.

This last fact can be derived with Ghouila-Houri's characterization (4.4) (ii). Choose a set of rows of C', i.e., choose a subcollection $\mathcal{G}$ of $\mathcal{F}$. Define, for each U in $\mathcal{G}$, the "height" $h(U)$ of U as the number of sets T in $\mathcal{G}$ with $T \supseteq U$. Now split $\mathcal{G}$ into $\mathcal{G}_{\text{odd}}$ and $\mathcal{G}_{\text{even}}$, according as $h(U)$ is odd or even. One easily derives from the laminarity of $\mathcal{G}$ that for any arc a of D, the number of sets in $\mathcal{G}_{\text{odd}}$ entered by a, and the number of sets in $\mathcal{G}_{\text{even}}$ entered by a, differ by at most 1. Therefore, we can split the rows corresponding to $\mathcal{G}$ into two classes fulfilling Ghouila-Houri's criterion. So C' is totally unimodular.

So it suffices to find a laminar subcollection $\mathcal{F}$ or $\mathcal{H}$ so that the corresponding matrix C' satisfies (5.9). This can be done as follows. We may assume that all components of l are nonnegative. (If some component is negative, the maximum in (5.8) is infinite.) Choose a vector y that attains the maximum in (5.8), and for which

$$\sum_{U\in\mathcal{H}} y_U \cdot |U| \cdot |V\setminus U| \tag{5.10}$$

is as small as possible. Such a vector y exists by compactness arguments.

Define

$$\mathcal{F} := \{U \mid y_U > 0\}\,. \tag{5.11}$$

Then $\mathcal{F}$ is laminar. To see this, suppose there are $T, U \in \mathcal{F}$ with $T \not\subseteq U \not\subseteq T$ and $T \cap U \neq \emptyset$. Let $\varepsilon := \min\{y_T, y_U\} > 0$. Next reset:

$$\begin{aligned} y_T &:= y_T - \varepsilon\,, & y_{T\cap U} &:= y_{T\cap U} + \varepsilon\,,\\ y_U &:= y_U - \varepsilon\,, & y_{T\cup U} &:= y_{T\cup U} + \varepsilon\,, \end{aligned} \tag{5.12}$$

while y does not change in the other coordinates. By this resetting, $y^{\mathrm{T}}C$ does not increase in any coordinate (since $\varepsilon \cdot \chi^{\delta^-(T)} + \varepsilon \cdot \chi^{\delta^-(U)} \geqslant \varepsilon \cdot \chi^{\delta^-(T\cap U)} + \varepsilon \cdot \chi^{\delta^-(T\cup U)}$), while $y^{\mathrm{T}}\mathbf{1}$ does not change. However, the sum (5.10) did decrease, contradicting the minimality of (5.10). This shows that $\mathcal{F}$ is laminar.

We finally show that (5.9) holds. The inequality $\leqslant$ is trivial, since C' is a submatrix of C. The inequality $\geqslant$ follows from the fact that the vector y above attains the second maximum in (5.9), while y has 0's in the positions corresponding to rows of C not in C'. □

A direct consequence is that the *r-arborescence polytope* of $D = (V, A)$ (being the convex hull of the incidence vectors of r-arborescences) is determined by

$$\begin{aligned} &0 \leqslant x_a \leqslant 1\,, \quad a \in A\,,\\ &\sum_{a\in\delta^-(U)} x_a \geqslant 1\,, \quad \emptyset \neq U \subseteq V\setminus\{r\}\,. \end{aligned} \tag{5.13}$$

This is a result of Edmonds (1967). It follows, with the ellipsoid method, that a minimum-length r-arborescence can be found in polynomial time if and only if we can test (5.13) in polynomial time. This last is indeed possible: given $x \in \mathbb{Q}^A$, we first test if $0 \leqslant x_a \leqslant 1$ for each arc a; if $x_a < 0$ or $x_a > 1$ for some a, we have a separating hyperplane. Otherwise, consider x as a capacity function on the arcs of D, and find an r-cut C of minimum capacity (with an adaptation of Ford and Fulkerson's algorithm): if C has capacity at least 1, then (5.13) is satisfied; otherwise, C yields a hyperplane separating x from the polyhedron determined by (5.13).

For a characterization of the facets of the r-arborescence polytope, see Held and Karp (1970) and Giles (1975, 1978).

One similarly shows that for any directed graph $D=(V,A)$, the following system, in $x \in \mathbb{R}^A$, is TDI:

$$\begin{aligned} &x_a \geq 0\,, \quad a \in A\,, \\ &\sum_{a \in \delta^-(U)} x_a \geq 1\,, \quad \emptyset \neq U \subseteq V,\ \delta^+(U)=\emptyset\,, \end{aligned} \tag{5.14}$$

which is a result of Lucchesi and Younger (1978). It is equivalent to:

Lucchesi–Younger Theorem 5.15. *The minimum size of a directed-cut covering in a digraph $D=(V,A)$ is equal to the maximum number of pairwise disjoint directed cuts.*

Here a *directed cut* is a set of arcs of the form $\delta^-(U)$ with $\emptyset \neq U \neq V$, $\delta^+(U)=\emptyset$. A *directed-cut covering* is a set of arcs intersecting each directed cut, or equivalently, a set of arcs whose contraction makes the digraph strongly connected.

Note that the Lucchesi–Younger Theorem is of a self-refining nature: it implies that for any "length" function $l: A \to \mathbb{Z}_+$, the minimum length of a directed-cut covering is equal to the maximum number t of directed cuts $C_1, \dots, C_t$ (repetition allowed), so that no arc a is in more than $l(a)$ of these cuts. [To derive this from Theorem 5.15, replace each arc a by a directed path of length $l(a)$.] In this weighted form, the Lucchesi–Younger Theorem is easily seen to be equivalent to the total dual integrality of (5.14).

Application 5.16 (*Polymatroid intersection*). Let S be a finite set. A function $f: \mathcal{P}(S) \to \mathbb{R}$ is called *submodular* if

$$f(T)+f(U) \geq f(T \cap U)+f(T \cup U) \quad \text{for all } T, U \subseteq S\,. \tag{5.17}$$

There are several examples of submodular functions. For example, the rank function of any matroid is submodular (see chapters 9 and 11).

Let f_1, f_2 be two submodular functions on S, and consider the following system in the variable $x \in \mathbb{R}^S$:

$$\begin{aligned} &\text{(i)} \quad x_s \geq 0\,, \quad s \in S\,, \\ &\text{(ii)} \quad \sum_{s \in U} x_s \leq f_1(U)\,, \quad U \subseteq S\,, \\ &\text{(iii)} \quad \sum_{s \in U} x_s \leq f_2(U)\,, \quad U \subseteq S\,. \end{aligned} \tag{5.18}$$

Edmonds (1970, 1979) proved:

Theorem 5.19. *System* (5.18) *is TDI.*

Proof. The proof is similar to that of Theorem 5.7. Let $c: S \rightarrow \mathbb{Z}$, and consider the LP problem dual to maximizing $c^{\mathrm{T}}x$ over (5.18):

$$\min\left\{\sum_{U \subseteq S} y_U f_1(U) + \sum_{U \subseteq S} z_U f_2(U) \,\middle|\, y, z \in \mathbb{R}_+^{\mathcal{P}(S)};\ \sum_{U \subseteq S} (y_U + z_U)\chi^U \geq c\right\}. \tag{5.20}$$

We show that this minimum has an integral optimum solution, by a version of the "uncrossing" technique. Let y, z attain this minimum, so that

$$\sum_{U \subseteq S} (y_U + z_U) \cdot |U| \cdot |S \backslash U| \tag{5.21}$$

is as small as possible. Let

$$\mathcal{F} := \{U \subseteq S \mid y_U > 0\}. \tag{5.22}$$

We show that $\mathcal{F}$ forms a chain with respect to inclusion. Suppose not. Let T, $U \in \mathcal{F}$ with $T \not\subseteq U \not\subseteq T$. Let $\varepsilon := \min\{y_T, y_U\} > 0$. Next reset as in (5.12). Again, the modified y forms, with the original z, an optimum solution of (5.20) [since $\chi^T + \chi^U = \chi^{T \cap U} + \chi^{T \cup U}$ and $f_1(T) + f_1(U) \geq f_1(T \cap U) + f_1(T \cup U)$]. However, (5.21) did decrease, contradicting its minimality. This shows that $\mathcal{F}$ forms a chain. Similarly,

$$\mathcal{G} := \{U \subseteq S \mid z_U > 0\} \tag{5.23}$$

forms a chain.

Now (5.20) is equal to

$$\min\left\{\sum_{U \in \mathcal{F}} y_U f_1(U) + \sum_{U \in \mathcal{G}} z_U f_2(U) \,\middle|\, y \in \mathbb{R}_+^{\mathcal{F}},\ z \in \mathbb{R}_+^{\mathcal{G}}; \right.$$
$$\left.\sum_{U \in \mathcal{F}} y_U \chi^U + \sum_{U \in \mathcal{G}} z_U \chi^U \geq c\right\} \tag{5.24}$$

since y, z attain (5.20), using (5.22) and (5.23).

The constraint matrix in (5.24) is totally unimodular, as can be derived easily with Ghouila-Houri's criterion (4.4) (ii). Hence (5.24) has an integral optimum solution y, z. By extending y, z with 0-components, we obtain an integral optimum solution of (5.20). □

This result has several corollaries, as we shall see. If f_1 and f_2 are integer-valued submodular functions, then the total dual integrality of (5.18) implies that (5.18) determines an integral polyhedron. In particular, let f_1 and f_2 be the rank functions of two matroids $(S, \mathcal{I}_1)$ and $(S, \mathcal{I}_2)$. Then the following result of Edmonds (1970) follows.

Corollary 5.25. *The polytope* $\mathrm{conv}\{\chi^I \mid I \in \mathcal{I}_1 \cap \mathcal{I}_2\}$ *is determined by* (5.18).

Proof. Note that an integral vector satisfies (5.18) iff it is equal to χ^I for some I in $\mathcal{I}_1 \cap \mathcal{I}_2$. □

A special case is that if we have one matroid $(S, \mathcal{I})$, with rank function, say, f, then its *independence polytope* ($=\mathrm{conv.}\{\chi^I \mid I \in \mathcal{I}\}$) is determined by $x_s \geq 0$, $s \in S$; $\sum_{s \in U} x_s \leq f(U)$, $U \subseteq S$ (Edmonds 1971). So Corollary 5.25 concerns the intersection of two independence polytopes. The facets of independence polytopes, and of the intersection of two of them, are described by Giles (1975). Hausmann and Korte (1978) characterized adjacency on the independence polytope. See also Edmonds (1979) and Cunningham (1984).

Another direct consequence for matroids is:

Edmonds' Matroid Intersection Theorem 5.26. *The maximum size of a common independent set of two matroids* $(S, \mathcal{I}_1)$ *and* $(S, \mathcal{I}_2)$ *is equal to* $\min_{U \subseteq S} (f_1(U) + f_2(S \backslash U))$, *where* f_1 *and* f_2 *are the rank functions of these matroids.*

Proof. By Corollary 5.25, the maximum size of a common independent set is equal to $\max\{\mathbf{1}^{\mathrm{T}} x \mid x \text{ satisfies } (5.18)\}$, and hence, by the total dual integrality of (5.18), to

$$\min\left\{ \sum_{U \subseteq S} (y_U f_1(U) + z_U f_2(U)) \mid y, z \in \mathbb{Z}_+^{\mathcal{P}(S)};\ \sum_{U \subseteq S} (y_U + z_U)\chi^U \geq \mathbf{1} \right\}.$$

It is not difficult (using the nonnegativity, the monotonicity and the submodularity of f_1 and f_2) to derive that this last minimum is equal to the minimum in Theorem 5.26. □

For more consequences of Theorem 5.19, we refer to chapter 11.

The proofs of Theorems 5.7 and 5.19 given above are examples of a general proof technique for total dual integrality studied by Edmonds and Giles (1977). First show that there exists an optimum dual solution whose nonzero components correspond to a "nice" colelction of sets (e.g., laminar, a chain, "cross-free"). Next prove that such nice collections yield a restricted linear program with totally unimodular constraint matrix. Finally, appeal to Hoffman and Kruskal's Theorem to deduce the existence of an integral optimum dual solution for the restricted, and hence for the original, problem.

We now illustrate how total dual integrality helps in showing one of the pioneering successes of polyhedral combinatorics, the characterization of the matching polytope by Edmonds (1965). For the basic theory on matchings we refer to chapter 3.

Definition 5.27. The *matching polytope* of an undirected graph $G = (V, E)$ is the polytope $\operatorname{conv}\{\chi^M \mid M \text{ matching}\}$ in $\mathbb{R}^E$. Edmonds showed that this polytope is equal to the set of all vectors x in $\mathbb{R}^E$ satisfying

$$\begin{aligned}
&\text{(i)} \quad x_e \geqslant 0\,, \quad e \in E\,,\\
&\text{(ii)} \quad \sum_{e \ni v} x_e \leqslant 1\,, \quad v \in V\,,\\
&\text{(iii)} \quad \sum_{e \subseteq U} x_e \leqslant \lfloor \tfrac{1}{2}|U| \rfloor\,, \quad U \subseteq V\,.
\end{aligned} \tag{5.28}$$

Since the integral vectors satisfying (5.28) are exactly the incidence vectors χ^M of matchings M, it suffices to show that (5.28) determines an integral polyhedron. In fact, Cunningham and Marsh (1978) showed:

Theorem 5.29. *System* (5.28) *is TDI.*

This implies that for each $w \colon E \to \mathbb{Z}$, both optima in the LP-duality equation

$$\begin{aligned}
&\max\{w^{\mathrm{T}}x \mid x \text{ satisfies } (5.28)\}\\
&= \min\Big\{\sum_{v \in V} y_v + \sum_{U \subseteq V} z_U \lfloor \tfrac{1}{2}|U| \rfloor \;\Big|\; y \in \mathbb{R}^V_+,\, z \in \mathbb{R}^{\mathcal{P}(V)}_+;\\
&\qquad \forall e \in E\colon \sum_{v \in e} y_v + \sum_{U \supseteq e} z_U \geqslant w_e\Big\}
\end{aligned} \tag{5.30}$$

are attained by integral optimum solutions. It means: for each undirected graph $G = (V, E)$ and for each "weight" function $w \colon E \to \mathbb{Z}$

$$\begin{aligned}
&\max\{w(M) \mid M \text{ matching}\}\\
&= \min\Big\{\sum_{v \in V} y_v + \sum_{U \subseteq V} z_U \lfloor \tfrac{1}{2}|U| \rfloor \;\Big|\; y \in \mathbb{Z}^V_+,\, z \in \mathbb{Z}^{\mathcal{P}(V)}_+;\\
&\qquad \forall e \in E\colon \sum_{v \in e} y_v + \sum_{U \supseteq e} z_U \geqslant w_e\Big\}
\end{aligned} \tag{5.31}$$

Here $w(E') := \sum_{e \in E'} w_e$ for any subset E' of E. [Note that (5.31) contains the Tutte–Berge formula as special case, by taking $w = \mathbf{1}$.]

Proof of Theorem 5.29. We may assume that w is nonnegative, since replacing any negative component of w by 0 does not change any optimum in (5.31).

For any w, let ν_w denote the left-hand term in (5.31). It suffices to show that ν_w is not less than the right-hand term in (5.31) (since $\leqslant$ is trivial). Suppose (5.31) does not hold, and suppose we have chosen $G = (V, E)$ and $w \colon E \to \mathbb{Z}_+$ so that $|V| + |E| + w(E)$ is as small as possible. Then G is connected (otherwise, one of the components of G will form a smaller counterexample) and $w_e \geqslant 1$ for each edge e (otherwise we could delete e). Now there are two cases.

Case 1. There exists a vertex v covered by every maximum-weighted matching. In this case, let $w' \in \mathbb{Z}_+^E$ arise from w by decreasing the weights of edges incident to v by 1. Then $\nu_{w'} = \nu_w - 1$. Since $w'(E) < w(E)$, (5.31) holds for w'. Increasing component y_v of the optimal y for w' by 1, shows (5.31) for w.

Case 2. No vertex is covered by every maximum-weighted matching. Now let w' arise from w by decreasing all weights by 1. We show that $\nu_w \geq \nu_{w'} + \lfloor \frac{1}{2}|V| \rfloor$. This will imply (5.31) for w: since $w'(E) < w(E)$, (5.31) holds for w'. Increasing component z_V of the optimal z for w' by 1, shows (5.31) for w.

Assume $\nu_w < \nu_{w'} + \lfloor \frac{1}{2}|V| \rfloor$, and let M be a matching with $\nu_{w'} = w'(M)$, such that $w(M)$ is as large as possible. Then M leaves at least two vertices in V uncovered, since otherwise $w(M) = w'(M) + \lfloor \frac{1}{2}|V| \rfloor$, implying $\nu_w \geq w(M) = w'(M) + \lfloor \frac{1}{2}|V| \rfloor = \nu_{w'} + \lfloor \frac{1}{2}|V| \rfloor$.

Let u and v be not covered by M, and suppose we have chosen M, u and v so that the distance $d(u, v)$ in G is as small as possible. Then $d(u, v) > 1$, since otherwise augmenting M by $\{u, v\}$ would increase $w(M)$. Let t be an internal vertex of a shortest path between u and v. Let M' be a matching with $w(M') = \nu_w$ not covering t.

Now $M \,\Delta\, M'$ is a disjoint union of paths and circuits. Let P be the set of edges of the component of $M \,\Delta\, M'$ containing t. Then P forms a path starting in t and not covering both u and v (as t, u and v each have degree at most 1 in $M \,\Delta\, M'$). Say P does not cover u. Now the symmetric difference $M \,\Delta\, P$ is a matching with $|M \,\Delta\, P| \leq |M|$, and therefore

$$\begin{aligned} w'(M \,\Delta\, P) - w'(M) &= w(M \,\Delta\, P) - |M \,\Delta\, P| - w(M) + |M| \\ &\geq w(M \,\Delta\, P) - w(M) = w(M') - w(M' \,\Delta\, P) \geq 0 \,. \end{aligned} \tag{5.32}$$

Hence $\nu_{w'} = w'(M \,\Delta\, P)$ and $w(M \,\Delta\, P) \geq w(M)$. However, $M \,\Delta\, P$ does not cover t and u, and $d(u, t) < d(u, v)$, contradicting our choice of M, u, and v. □

So (5.28) is TDI. A consequence is the following fundamental result of Edmonds (1965).

Edmonds' Matching Polyhedron Theorem 5.33. *The matching polytope of a graph is equal to the polyhedron determined by* (5.28).

In fact, Edmonds found Theorem 5.33 as a by-product of a polynomial-time algorithm for finding a maximum-weighted matching. In turn, with the ellipsoid method, Padberg and Rao (1982) showed that Theorem 5.33 yields a polynomial-time algorithm finding a maximum-weighted matching, see (5.37) below.

5.34. A consequence of Theorem 5.33 is a characterization of the *perfect matching polytope* of a graph $G = (V, E)$, which is the polytope $\text{conv}\{\chi^M \mid M$ perfect matching$\}$ in $\mathbb{R}^E$. This polytope clearly is a face of the matching polytope of G (or is empty), viz. the intersection of the matching polytope with the (supporting) hyperplane $\{x \in \mathbb{R}^E \mid \sum_{e \in E} x_e = \frac{1}{2}|V|\}$. It follows that the perfect matching polytope is determined by the following inequalities:

$$
\begin{array}{lll}
\text{(i)} & x_e \geqslant 0\,, \quad e \in E\,, & \\
\text{(ii)} & \sum_{e \ni v} x_e = 1\,, \quad v \in V\,, & (5.35) \\
\text{(iii)} & \sum_{e \in \delta(U)} x_e \geqslant 1\,, \quad U \subseteq V,\ |U| \text{ odd}\,. &
\end{array}
$$

(Note that (ii) and (iii) imply (5.28) (iii).)

From the description (5.35) of the perfect matching polytope one can derive with the ellipsoid method a polynomial-time algorithm for finding a maximum-weighted perfect matching (and through this a maximum-weighted matching). It amounts to showing that it can be tested in polynomial time whether a vector x satisfies (5.35). Padberg and Rao (1982) showed that this can be done as follows.

For a given $x \in \mathbb{Q}^E$ we must test if x satisfies (5.35). The inequalities in (i) and (ii) can be checked one by one. If one of them is not satisfied, it gives us a separating hyperplane. So we may assume that (i) and (ii) are satisfied. If $|V|$ is odd, then clearly (iii) is not satisfied for $U := V$. So we may assume that $|V|$ is even. We cannot check the constraints in (iii) one by one in polynomial time, simply because there are exponentially many of them. Yet, there is a polynomial-time method of checking time. First, note that from Ford and Fulkerson's max-flow min-cut algorithm we can easily derive a polynomial-time algorithm having the following as input and output:

Input: Subset W of V.
Output: Subset T of V such that $W \cap T \neq \emptyset \neq W \backslash T$ and such that $x(\delta(T))$ is as small as possible. (5.36)

Here $x(E') := \sum_{e \in E'} x_e$ for any subset E' of E. We next describe recursively an algorithm with the following input and output specification:

Input: Subset W of V with $|W|$ even.
Output: Subset U of V such that $|W \cap U|$ is odd and such that $x(\delta(U))$ is as small as possible. (5.37)

First, we find with algorithm (5.36) a subset T of V with $W \cap T \neq \emptyset \neq W \backslash T$ and with $x(\delta(T))$ minimal. If $|W \cap T|$ is odd, we are done. If $|W \cap T|$ is even, call, recursively, the algorithm (5.37) for the inputs $W \cap T$ and $W \cap \bar{T}$, respectively, where $\bar{T} := V \backslash T$. Let it yield a subset U' of V such that $|W \cap T \cap U'|$ is odd and $x(\delta(U'))$ is minimal, and a subset U'' of V such that $|W \cap \bar{T} \cap U''|$ is odd and $x(\delta(U''))$ is minimal. Without loss of generality, $W \cap \bar{T} \not\subseteq U'$ (otherwise replace U' by $V \backslash U'$), and $W \cap T \not\subseteq U''$ (otherwise replace U'' by $V \backslash U''$).

We claim that if $x(\delta(T \cap U')) < x(\delta(\bar{T} \cap U''))$, then $U := T \cap U'$ is output of (5.37) for input W, and otherwise $U := \bar{T} \cap U''$. To see that this output is justified suppose to the contrary that there exists a subset Y of V such that $|W \cap Y|$ is odd, and $x(\delta(Y)) < x(\delta(T \cap U'))$ and $x(\delta(Y)) < x(\delta(\bar{T} \cap U''))$. Then either $|W \cap Y \cap T|$ is odd or $|W \cap Y \cap \bar{T}|$ is odd.

Case 1. $|W \cap Y \cap T|$ is odd. Then $x(\delta(Y)) \geq x(\delta(U'))$, since U' is output of (5.37) for input $W \cap T$. Moreover, $x(\delta(T \cup U') \geq x(\delta(T))$, since T is output of (5.36) for input W, and since $W \cap (T \cup U') \neq \emptyset \neq W \backslash (T \cup U')$. Therefore, we have the following contradiction:

$$\begin{aligned} x(\delta(Y)) &\geq x(\delta(U')) \geq x(\delta(T \cap U')) + x(\delta(T \cup U')) - x(\delta(T)) \\ &\geq x(\delta(T \cap U')) > x(\delta(Y)) \end{aligned} \tag{5.38}$$

[the second inequality follows since $x(\delta(A)) + x(\delta(B)) \geq x(\delta(A \cap B)) + x(\delta(A \cup B))$ for all $A, B \subseteq V$].

Case 2. $|W \cap Y \cap \bar{T}|$ is odd: similar.

Given the polynomial speed of the algorithm for (5.36), it is not difficult to see that the algorithm described for (5.37) is also polynomial-time. As a consequence, we can test (5.35) (iii) in polynomial time.

Further notes on TDI: for a deep characterization of certain TDI systems, see Seymour (1977). For an application of TDI to non-optimizational combinatorics (viz. Nash–Williams Orientation Theorem), see Frank (1980), and Frank and Tardos (1984).

6. Blocking polyhedra

Another useful technique in polyhedral combinatorics is a variant of the classical polarity in Euclidean space, viz. the blocking relation between polyhedra. It was introduced by Fulkerson (1970a, 1971), who noticed its importance to combinatorics and optimization. Often, with the theory of blocking polyhedra, one polyhedral characterization (or min–max relation) can be derived from another, and conversely.

The basic idea is the following result. Let $c_1, \ldots, c_m, d_1, \ldots, d_t \in \mathbb{R}^n_+$ satisfy

$$\operatorname{conv}\{c_1, \ldots, c_m\} + \mathbb{R}^n_+ = \{x \in \mathbb{R}^n_+ \mid d_j^{\mathrm{T}} x \geq 1 \text{ for } j = 1, \ldots, t\}\,. \tag{6.1}$$

Then the same holds after interchanging the c_i and d_j:

$$\operatorname{conv}\{d_1, \ldots, d_t\} + \mathbb{R}^n_+ = \{x \in \mathbb{R}^n_+ \mid c_i^{\mathrm{T}} x \geq 1 \text{ for } i = 1, \ldots, m\}. \tag{6.2}$$

In a sense, in (6.2) the ideas of "vertex" and "facet" are interchanged as compared with (6.1). The proof is a simple application of Farkas's Lemma.

Theorem 6.3. *For any* $c_1, \ldots, c_m, d_1, \ldots, d_t \in \mathbb{R}^n_+$, (6.1) *holds if and only if* (6.2) *holds.*

Proof. Suppose (6.1) holds. Then $\subseteq$ in (6.2) is direct, since $c_i^{\mathrm{T}} d_j \geq 1$ for all i, j, as the c_i belong to the right-hand side in (6.1), and since $c \geq 0$.

To show $\supseteq$ in (6.2), suppose $x \not\in \operatorname{conv}\{d_1, \ldots, d_t\} + \mathbb{R}^n_+$. Then there exists a separating hyperplane, i.e., there is a vector y such that

$$y^{\mathrm{T}}x > \min\{y^{\mathrm{T}}z \mid z \in \operatorname{conv}\{d_1, \ldots, d_t\} + \mathbb{R}^n_+\} . \tag{6.4}$$

We may assume $t \geqslant 1$ [since if $t = 0$, then (6.1) gives that $0 \in \{c_1, \ldots, c_m\}$, and therefore x does not belong to the right-hand side of (6.2)]. By scaling y, we can assume that the minimum in (6.4) is 1. Therefore, y belongs to the right-hand side of (6.1), and therefore to the left-hand side. So $y \geqslant \lambda_1 c_1 + \cdots + \lambda_m c_m$ for certain $\lambda_1, \ldots, \lambda_m \geqslant 0$ with $\lambda_1 + \cdots + \lambda_m = 1$. Since $y^{\mathrm{T}}x < 1$, it follows that $c_i^{\mathrm{T}}x < 1$ for at least one i. Hence x does not belong to the right-hand side of (6.2).

This shows (6.1) $\Rightarrow$ (6.2). The reverse implication follows by symmetry. □

This theorem has the following consequences. For any $X \subseteq \mathbb{R}^n$, we define the *blocker* $B(X)$ of X by:

$$B(X) := \{x \in \mathbb{R}^n_+ \mid y^{\mathrm{T}}x \geqslant 1 \text{ for each } y \text{ in } X\} . \tag{6.5}$$

Clearly, for $c_1, \ldots, c_m \in \mathbb{R}^n_+$, if P is the polyhedron

$$P := \operatorname{conv}\{c_1, \ldots, c_m\} + \mathbb{R}^n_+ , \tag{6.6}$$

then

$$B(P) = \{x \in \mathbb{R}^n_+ \mid c_i^{\mathrm{T}}x \geqslant 1 \text{ for } i = 1, \ldots, m\} . \tag{6.7}$$

So $B(P)$ is also a polyhedron, called the *blocking polyhedron* of P. If $R = B(P)$, the pair P, R is called a *blocking pair* of polyhedra. By the following direct corollary of Theorem 6.3, this is a symmetric relation.

Corollary 6.8. *For any polyhedron of type* (6.6), $B(B(P)) = P$.

So both (6.1) and (6.2) are equivalent to:

the pair $\operatorname{conv}\{c_1, \ldots, c_m\} + \mathbb{R}^n_+$ and $\operatorname{conv}\{d_1, \ldots, d_t\} + \mathbb{R}^n_+$ forms a blocking pair of polyhedra. (6.9)

The following corollary shows the equivalence of certain min–max relations.

Corollary 6.10. *Let* $c_1, \ldots, c_m, d_1, \ldots, d_t \in \mathbb{R}^n_+$. *Then the following are equivalent*:

(i) *for each* $l \in \mathbb{R}^n$: $\min\{l^{\mathrm{T}}c_1, \ldots, l^{\mathrm{T}}c_m\}$

$$= \max\Big\{\lambda_1 + \cdots + \lambda_t \mid \lambda_1, \ldots, \lambda_t \in \mathbb{R}_+ ; \sum_j \lambda_j d_j \leqslant l\Big\} ; \tag{6.11}$$

(ii) *for each* $w \in \mathbb{R}^n_+$: $\min\{w^{\mathrm{T}}d_1, \ldots, w^{\mathrm{T}}d_t\}$

$$= \max\left\{\mu_1 + \cdots + \mu_m \mid \mu_1, \ldots, \mu_m \in \mathbb{R}_+ ; \sum_i \mu_i c_i \leq w\right\} . \tag{6.12}$$

Proof. By LP duality, the maximum in (6.11) is equal to min $\{l^{\mathrm{T}}x \mid x \in \mathbb{R}^n_+ ; d_j^{\mathrm{T}}x \geq 1$ for $j = 1, \ldots, t\}$. Hence, (6.11) is equivalent to (6.1). Similarly, (6.12) is equivalent to (6.2). Therefore, Theorem 6.3 implies Corollary 6.10. □

Note that by continuity, in (6.11) we may restrict l to rational, and hence to integral vectors, without changing the condition. Similarly for (6.12). This is sometimes useful when showing one of them by induction.

A symmetric characterization of the blocking relation is the "length–width inequality" given by Lehman (1965):

Lehman's Length–Width inequality 6.13. *Let* $c_1, \ldots, c_m, d_1, \ldots, d_t \in \mathbb{R}^n_+$. *Then* (6.1) [*equivalently* (6.2), (6.11), *or* (6.12)] *holds if and only if*

$$\begin{aligned} &\text{(i)} \quad d_j^{\mathrm{T}} c_i \geq 1 \textit{ for all } i = 1, \ldots, m \textit{ and } j = 1, \ldots, t; \\ &\text{(ii)} \quad \min\{l^{\mathrm{T}}c_1, \ldots, l^{\mathrm{T}}c_m\} \cdot \min\{w^{\mathrm{T}}d_1, \ldots, w^{\mathrm{T}}d_t\} \leq l^{\mathrm{T}}w \textit{ for all } l, w \in \mathbb{Z}^n_+ . \end{aligned} \tag{6.14}$$

Proof. Suppose (6.14) holds. We derive (6.11). Let $l \in \mathbb{R}^n_+$. By LP duality, the maximum in (6.11) is equal to $\min\{l^{\mathrm{T}}x \mid x \in \mathbb{R}^n_+ ; d_j^{\mathrm{T}}x \geq 1$ for $j = 1, \ldots, t\}$. Let this minimum be attained by vector w. Then by (6.14)

$$l^{\mathrm{T}}w \geq (\min_i l^{\mathrm{T}}c_i)(\min_j w^{\mathrm{T}}d_j) \geq \min_i l^{\mathrm{T}}c_i \geq l^{\mathrm{T}}w .$$

So the minimum in (6.11) is equal to $l^{\mathrm{T}}w$.

Next, suppose (6.1) holds. Then (6.11) and (6.12) hold. Now (6.14) (i) follows by taking $l = d_j$ in (6.11). To show (6.14) (ii), let $\lambda_1, \ldots, \lambda_t, \mu_1, \ldots, \mu_m$ attain the maxima in (6.11) and (6.12). Then

$$\Big(\sum_j \lambda_j\Big)\Big(\sum_i \mu_i\Big) = \sum_j \sum_i \lambda_j \mu_i \leq \sum_j \sum_i \lambda_j \mu_i d_j^{\mathrm{T}} c_i = \Big(\sum_j \lambda_j d_j\Big)^{\mathrm{T}} \Big(\sum_i \mu_i c_i\Big)$$
$$\leq l^{\mathrm{T}} w .$$

This implies (6.14) (ii). □

It follows from the ellipsoid method that if $c_1, \ldots, c_m, d_1, \ldots, d_t \in \mathbb{R}^n_+$ satisfy (6.1) [equivalently, (6.2), (6.11), or (6.12)], then

for each $l \in \mathbb{R}^n_+$: $\min\{l^{\mathrm{T}}c_1, \ldots, l^{\mathrm{T}}c_m\}$ can be found in polynomial time if and only if
for each $w \in \mathbb{R}^n_+$: $\min\{w^{\mathrm{T}}d_1, \ldots, w^{\mathrm{T}}d_t\}$ can be found in polynomial time. (6.17)

This is particularly interesting if t or m is exponentially large (cf. the applications below).

For more on blocking (and anti-blocking) polyhedra, see Aráoz (1973), Aráoz et al. (1983), Bland (1978), Griffin (1977), Griffin et al. (1982), Huang and Trotter (1980), and Johnson (1978).

Application 6.18 (*Shortest paths and network flows*). The theory of blocking polyhedra yields an illustrative short proof of the Max-Flow-Min-Cut Theorem. Let $D=(V,A)$ be a directed graph, and let $r, s \in V$. Let $c_1,\ldots,c_m \in \mathbb{R}^A_+$ be the incidence vectors of the r–s-paths in D. Similarly, let $d_1,\ldots,d_t \in \mathbb{R}^A_+$ be the incidence vectors of the r–s-cuts.

Considering a given function $l\colon A \to \mathbb{Z}_+$ as a "length" function, one easily verifies: the minimum length of an r–s-path is equal to the maximum number of r–s-cuts (repetitition allowed) so that no arc a is in more than $l(a)$ of these cuts. [Indeed, the inequality min $\geqslant$ max is easy. To see the reverse inequality, let p be the minimum length of an r–s-path. For $i=1,\ldots,p$, let

$$V_i := \{v \in V \mid \text{the shortest } r\text{–}v\text{-path has length at least } i\}.$$

Then $\delta^-(V_1),\ldots,\delta^-(V_p)$ are r–s-cuts as required.] This implies (6.11). Hence (6.12) holds, which is equivalent to the Max-Flow Min-Cut Theorem: the maximum amount of r–s-flow subject to a capacity function w is equal to the minimum capacity of an r–s-cut. (Note that $\sum_i \mu_i c_i$ is an r–s-flow.) In fact, there exists an integral optimum flow if the capacities are integer, but this fact does not seem to follow from the theory of blocking polyhedra.

The above implies that the polyhedra $\operatorname{conv}\{c_1,\ldots,c_m\}+\mathbb{R}^A_+$ and $\operatorname{conv}\{d_1,\ldots,d_t\}+\mathbb{R}^A_+$ form a blocking pair of polyhedra. By (6.17), the polynomial-time solvability of the minimum-capacitated cut problem is equivalent to that of the shortest-path problem; note that this latter problem is much easier.

Application 6.19 (r-*arborescence*). Let $D=(V,A)$ be a digraph and let $r \in V$. Let $c_1,\ldots,c_m$ be the incidence vectors of r-arborescences, and let $d_1,\ldots,d_t$ be the incidence vectors of r-cuts (cf. Application 5.6).

From (5.13) we know that (6.1) holds. Therefore, by Theorem 6.3, also (6.2) holds. It means that for any "capacity" function $w \in \mathbb{R}^A_+$, the minimum capacity of an r-cut is equal to the maximum value of $\mu_1+\cdots+\mu_k$ where $\mu_1,\ldots,\mu_k \geqslant 0$ are such that there exist r-arborescences $T_1,\ldots,T_k$ with the property that for each arc a, the sum of the μ_j for which $a \in T_j$ is at most c_a.

Hence the convex hull of the incidence vectors of sets containing an r-cut as a subset, is determined by the system (in $x \in \mathbb{R}^A$)

$$\begin{array}{lll} \text{(i)} & 0 \leqslant x_a \leqslant 1\,, & a \in A\,, \\ \text{(ii)} & \sum\limits_{a \in T} x_a \geqslant 1\,, & T\ r\text{-arborescence}\,. \end{array} \tag{6.20}$$

Edmonds (1973) in fact showed that (6.20) is TDI (again, this does not seem to follow from the theory of blocking polyhedra). It is equivalent to: the minimum size of an r-cut is equal to the maximum number of pairwise disjoint r-arborescences.

The theory of blocking polyhedra can also be applied to directed cuts and directed-cut covers (cf. Theorem 5.15). Again it follows that the convex hull of incidence vectors of sets containing a directed cut as a subset, is determined by (6.20), with "r-arborescence" replaced by "directed-cut cover". However, in this case the system is not TDI (cf. Schrijver 1980b, 1982, 1983a).

Similar arguments apply to T-joins and T-cuts.

7. Anti-blocking polyhedra

The theory of anti-blocking polyhedra, due to Fulkerson (1971, 1972), is to a large extent parallel to that of blocking polyhedra, and arises mostly by reversing inequality signs and by interchanging "min" and "max". We here restrict ourselves to listing results analogous to those given in section 6, the proofs being similar.

Let $c_1, \ldots, c_m$, $d_1, \ldots, d_t \in \mathbb{R}^n_+$ be such that $\dim(\langle c_1, \ldots, c_m \rangle) = \dim(\langle d_1, \ldots, d_t) = n$. Then the following are equivalent:

$$(\operatorname{conv}\{c_1, \ldots, c_m\} + \mathbb{R}^n_-) \cap \mathbb{R}^n_+ = \{x \in \mathbb{R}^n_+ \mid d_j^{\mathrm{T}} x \leqslant 1 \text{ for } j = 1, \ldots, t\}\,, \tag{7.1}$$

$$(\operatorname{conv}\{d_1, \ldots, d_t\} + \mathbb{R}^n_-) \cap \mathbb{R}^n_+ = \{x \in \mathbb{R}^n_+ \mid c_i^{\mathrm{T}} x \leqslant 1 \text{ for } i = 1, \ldots, m\}\,. \tag{7.2}$$

Define for any subset X of $\mathbb{R}^n$ the *anti-blocker* $A(X)$ of X by:

$$A(X) := \{x \in \mathbb{R}^n_+ \mid y^{\mathrm{T}} x \leqslant 1 \text{ for each } y \text{ in } X\}\,.$$

Clearly, if

$$P := (\operatorname{conv}\{c_1, \ldots, c_m\} + \mathbb{R}^n_-) \cap \mathbb{R}^n_+\,, \tag{7.4}$$

then

$$A(P) = \{x \in \mathbb{R}^n_+ \mid c_i^{\mathrm{T}} x \leqslant 1 \text{ for } i = 1, \ldots, m\}\,. \tag{7.5}$$

$A(P)$ is called the *anti-blocking polyhedron* of P. If $R = A(P)$, the pair P, R is called an *anti-blocking pair* of polyhedra. Again, this is a symmetric relation:

$$\text{For any polyhedron } P \text{ of type (7.4), } A(A(P)) = P\,. \tag{7.6}$$

Each of the following are equivalent among themselves and to (7.1) and (7.2):

(a) The pair $(\operatorname{conv}\{c_1, \ldots, c_m\} + \mathbb{R}^n_-) \cap \mathbb{R}^n_+$ and $(\operatorname{conv}\{d_1, \ldots, d_t\} + \mathbb{R}^n_-)$

$\cap \mathbb{R}^n_+$ forms an anti-blocking pair of polyhedra; (7.7)

(b) For each $l \in \mathbb{R}^n_+$: $\max\{l^{\mathrm{T}}c_1, \ldots, l^{\mathrm{T}}c_m\}$

$$= \min\left\{\lambda_1 + \cdots + \lambda_t \mid \lambda_1, \ldots, \lambda_t \in \mathbb{R}_+; \sum_j \lambda_j d_j \geq l\right\}; \tag{7.8}$$

(c) For each $w \in \mathbb{R}^n_+$: $\max\{w^{\mathrm{T}}d_1, \ldots, w^{\mathrm{T}}d_t\}$

$$= \min\left\{\mu_1 + \cdots + \mu_m \mid \mu_1, \ldots, \mu_m \in \mathbb{R}_+; \sum_i \mu_i c_i \geq w\right\}; \tag{7.9}$$

(d) (i) $d_j^{\mathrm{T}}c_i \leq 1$ for all $i = 1, \ldots, m$ and $j = 1, \ldots, t$,
(ii) $\max\{l^{\mathrm{T}}c_1, \ldots, l^{\mathrm{T}}c_m\} \cdot \max\{w^{\mathrm{T}}d_1, \ldots, w^{\mathrm{T}}d_t\} \geq l^{\mathrm{T}}w$ for all $l, w \in \mathbb{Z}^n_+$. (7.10)

This last characterization is again due to Lehman (1965).

Application 7.11 (*Perfect graphs*). The theory of anti-blocking polyhedra yields a proof of Lovász's Perfect Graph Theorem (cf. chapter 4). This line of proof was developed by Fulkerson (1970b, 1972, 1973), Lovász (1972), and Chvátal (1975).

Define for any graph $G = (V, E)$, the *stable-set polytope* STAB(G) of G as the convex hull of the incidence vectors of stable sets in G. Clearly, any vector x in the stable-set polytope satisfies

(i) $x_v \geq 0$, $v \in V$,
(ii) $\sum_{v \in K} x_v \leq 1$, $K \subseteq V$, K clique, (7.12)

since the incidence vector of any stable set satisfies (7.12). Note that the polytope determined by (7.12) is exactly $A(\mathrm{STAB}(\bar{G}))$. The circuit on five vertices shows that generally $A(\mathrm{STAB}(\bar{G}))$ can be larger than STAB(G). Chvátal (1975) showed that STAB(G) is exactly determined by (7.12) if and only if G is perfect. Anti-blocking then yields the Perfect Graph Theorem.

First observe the following. Let $Ax \leq \mathbf{1}$ denote the inequality system (7.12) (ii). So the rows of A are the incidence vectors of cliques. By definition, G is perfect if and only if the (dual) linear programs

$$\max\{w^{\mathrm{T}}x \mid x \geq 0, Ax \leq \mathbf{1}\} = \min\{y^{\mathrm{T}}\mathbf{1} \mid y \geq 0, y^{\mathrm{T}}A \geq w^{\mathrm{T}}\} \tag{7.13}$$

have integral optimum solutions, for each $\{0, 1\}$-vector w.

Chvátal's Theorem 7.14. *G is perfect if and only if its stable-set polytope is determined by* (7.12).

Proof. (I) First suppose G is perfect. For $w: V \to \mathbb{Z}_+$, let α_w denote the maximum weight of a stable set. To prove that the stable-set polytope is determined by

(7.12), it suffices to show that

$$\alpha_w = \max\{w^{\mathrm{T}}x \mid x \geq 0,\ Ax \leq \mathbf{1}\} \tag{7.15}$$

for each $w: V \to \mathbb{Z}_+$. This will be done by induction on $\sum_{v \in V} w_v$.

If w is a $\{0, 1\}$-vector, then (7.15) follows from the remark on (7.13). So we may assume that $w_u \geq 2$ for some vertex u. Let $e_u = 1$ and $e_v = 0$ if $v \neq u$. Replacing w by $w - e$ in (7.13) and (7.15) gives, by induction, a vector $y \geq 0$ so that $y^{\mathrm{T}}A \geq (w-e)^{\mathrm{T}}$ and $y^{\mathrm{T}}\mathbf{1} = \alpha_{w-e}$. Since $(w-e)_u \geq 1$, there is a clique K with $y_K > 0$ and $u \in K$. We may assume that $\chi^K \leq w - e$. Denote $a := \chi^K$.

Then $\alpha_{w-a} < \alpha_w$. For suppose $\alpha_{w-a} = \alpha_w$. Let S be any stable set with $\sum_{v \in S} (w-a)_v = \alpha_{w-a}$. Since $\alpha_{w-a} = \alpha_w$, $K \cap S = \emptyset$. On the other hand, since $w - a \leq w - e \leq w$, we know that $\sum_{v \in K} (w-e)_v = \alpha_{w-e}$ and hence, by complementary slackness, $|K \cap S| = 1$, which is a contradiction.

Therefore,

$$\begin{aligned}\alpha_w &= 1 + \alpha_{w-a} = 1 + \max\{(w-a)^{\mathrm{T}}x \mid x \geq 0,\ Ax \leq \mathbf{1}\}\\ &\geq \max\{w^{\mathrm{T}}x \mid x \geq 0,\ Ax \leq \mathbf{1}\}\ ,\end{aligned} \tag{7.16}$$

implying (7.15).

(II) Conversely, suppose that the stable-set polytope is determined by (7.12), i.e., that the maximum in (7.13) is attained by the incidence vector of a stable set, for each $w \in \mathbb{Z}_+^V$. To show that G is perfect it suffices to show that the minimum in (7.13) also has an integer optimum solution for each $\{0, 1\}$-valued w. This will be done by induction on $\sum_{v \in V} w_v$.

Let w be $\{0, 1\}$-valued, and let y be a, not necessarily integral, optimum solution for the minimum in (7.13). Let K be a clique with $y_K > 0$, and let $a = \chi^K$ (we may assume $a \leq w$). Then the common value of

$$\max\{(w-a)^{\mathrm{T}}x \mid x \geq 0,\ Ax \leq \mathbf{1}\} = \min\{y^{\mathrm{T}}\mathbf{1} \mid y \geq 0,\ y^{\mathrm{T}}A \geq (w-a)^{\mathrm{T}}\} \tag{7.17}$$

is less than the common value of (7.13), since by complementary slackness, each optimum solution x in (7.13) has $a^{\mathrm{T}}x = 1$. However, the values in (7.13) and (7.17) are integers (since by assumption, the maxima have integral optimum solutions). Hence they differ by exactly 1. Moreover, by induction the minimum in (7.17) has an integral optimum solution y. Increasing component y_K of y by 1, gives an integral optimum solution of (7.13). □

Equivalent to Theorem 7.14 is:

$$G \text{ is perfect} \iff \mathrm{STAB}(G) = A(\mathrm{STAB}(\bar{G}))\ . \tag{7.18}$$

Note that the stable-set polytope of G is determined by (7.12) if the stable-set polytope and the clique polytope of G form an anti-blocking pair of polyhedra. Here the *clique polytope* is the convex hull of the incidence vectors of cliques.

The theory of anti-blocking polyhedra then gives directly the Perfect Graph Theorem of Lovász (1972):

Lovász's Perfect Graph Theorem 7.19. *The complement of a perfect graph is perfect.*

Proof. If G is perfect, then $\mathrm{STAB}(G) = A(\mathrm{STAB}(\bar{G}))$. Hence $\mathrm{STAB}(\bar{G}) = A(A(\mathrm{STAB}(\bar{G}))) = A(\mathrm{STAB}(G))$. Therefore, $\bar{G}$ is perfect. □

By (7.14), with the ellipsoid method, a maximum-weighted stable set in a perfect graph G can be found in polynomial time if and only if a maximum-weighted clique in a perfect graph G can be found in polynomial time. Since the complement of a perfect graph is a perfect graph again, this would not give any reduction of one problem to another.

However, an alternative approach does give a polynomial-time algorithm to find a maximum-weighted stable set in a perfect graph (Grötschel et al. 1981, 1986, 1988). Let $G = (V, E)$ be a graph, with $V = \{1, \ldots, n\}$, say. *Consider the* collection $M(G)$ of all matrices $Y = (y_{ij})_{i,j=0}^{n}$ in $\mathbb{R}^{(n+1)\times(n+1)}$ satisfying

$$\begin{aligned}
&\text{(i)} \quad Y \text{ is symmetric and positive semi-definite;} \\
&\text{(ii)} \quad y_{00} = 1\,, \qquad y_{0i} = y_{ii}\,, \quad i = 1, \ldots, n\,; \qquad\qquad (7.20)\\
&\text{(iii)} \quad y_{ij} = 0 \text{ if } i \neq j,\ \{i, j\} \in E\,.
\end{aligned}$$

These conditions imply that $M(K)$ is a convex set (not necessarily a polytope).

Let $\mathrm{TH}(G)$ be the set of all vectors $x \in \mathbb{R}^n$ for which there exists a matrix Y in $M(G)$ so that $x_i = y_{ii}$ for $i = 1, \ldots, n$. So $\mathrm{TH}(G)$ is the projection of $M(G)$ on the diagonal coordinates [excluding the $(0, 0)$ coordinate].

Now $\mathrm{TH}(G)$ turns out to be an approximation of $\mathrm{STAB}(G)$, at least as good as $A(\mathrm{STAB}(\bar{G}))$, in the following sense:

Theorem 7.21. $\mathrm{STAB}(G) \subseteq \mathrm{TH}(G) \subseteq A(\mathrm{STAB}(\bar{G}))$.

Proof. The first inclusion follows from the fact that for each stable set $S \subseteq V$, the incidence vector χ^S belongs to $\mathrm{TH}(G)$, as it is the projection of the matrix Y in $M(G)$ defined by:

$$y_{ij} = \begin{cases} 1 & \text{if } i, j \in S \cup \{0\}\,, \\ 0 & \text{otherwise}\,. \end{cases} \tag{7.22}$$

To see the second inclusion, first note that trivially each vector in $\mathrm{TH}(G)$ is nonnegative (since the diagonal of a positive semi-definite matrix is nonnegative). It next suffices to show: if $x \in \mathrm{TH}(G)$ and u is the incidence vector of a stable set in $\bar{G}$, then $u^{\mathrm{T}}x \leq 1$. To prove this, let x be the projection of $Y \in M(G)$. Since Y is positive semi-definite we know:

$$(1 \quad -u^{\mathrm{T}})Y\binom{1}{-u} \geqslant 0\,. \tag{7.23}$$

As $y_{ij} = 0$ if $\{i, j\} \in E$, and as u is the incidence vector of a clique K in G, (7.23) implies

$$1 - 2\sum_{i\in K} y_{i0} + \sum_{i\in K} y_{ii} \geqslant 0\,. \tag{7.24}$$

Since $x_i = y_{i0} = y_{ii}$, this implies $u^{\mathrm{T}}x \leqslant 1$. □

Theorem 7.21 implies that if $\mathrm{STAB}(G) = A(\mathrm{STAB}(\bar{G}))$, i.e., if G is perfect then $\mathrm{STAB}(G) = \mathrm{TH}(G)$. Now any linear objective function $w^{\mathrm{T}}x$ can be maximized over $\mathrm{TH}(G)$ in polynomial time. This follows from the fact that any linear objective function can be maximized over $M(G)$ in polynomial time, since we can solve the separation problem over $M(G)$ in polynomial time. [The latter follows from the fact that we can test, for any given Y in $\mathbb{R}^{(n+1)\times(n+1)}$, the constraints in (7.20) in polynomial time, in such a way that we find a separating hyperplane (in the space $\mathbb{R}^{(n+1)\times(n+1)}$) if Y does not belong to $M(G)$.]

So as a consequence we have:

Theorem 7.25. *There exists a polynomial-time algorithm finding a maximum-weight stable set in any given perfect graph.*

By symmetry, the same holds for finding a maximum-weight clique in a perfect graph.

Application 7.26 (*Matchings and edge-colorings*). Let for any graph $G = (V, E)$, $P_{\mathrm{mat}}(G)$ denote the matching polytope of G. By scalar multiplication, we can normalize system (5.28) determining $P_{\mathrm{mat}}(G)$ to : $x \geqslant 0$, $Cx \leqslant \mathbf{1}$, for a certain matrix C (deleting the inequalities in (5.28) corresponding to $U \subseteq V$ with $|U| \leqslant 1$). The matching polytope is of type (7.4), and hence its anti-blocking polyhedron $A(P_{\mathrm{mat}}(G))$ is equal to $\{z \in \mathbb{R}_+^E \mid Dz \leqslant \mathbf{1}\}$, where the rows of D are the incidence vectors of all matchings in G. So by (7.8), taking $l = \mathbf{1}$:

$$\max\left\{\Delta(G),\ \max_{U\subseteq V, |U|>2} \frac{|\langle U\rangle|}{\lfloor \frac{1}{2}|U|\rfloor}\right\} = \min\{y^{\mathrm{T}}\mathbf{1} \mid y \geqslant 0,\ y^{\mathrm{T}}D \geqslant \mathbf{1}^{\mathrm{T}}\}\,. \tag{7.27}$$

Here $\langle U\rangle$ denotes the collection of all edges contained in U.

The minimum in (7.27) can be interpreted as the *fractional edge-coloring number* $\gamma^*(G)$ of G. If the minimum is attained by an integral optimum solution y, it is equal to the edge-coloring number $\gamma(G)$ of G, since

$$\gamma(G) = \min\{y^{\mathrm{T}}\mathbf{1} \mid y \geqslant 0,\ y^{\mathrm{T}}D \geqslant \mathbf{1}^{\mathrm{T}},\ y \text{ integral}\}\,. \tag{7.28}$$

By Vizing's Theorem, $\gamma(G) = \Delta(G)$ or $\gamma(G) = \Delta(G) + 1$ if G is a simple graph. If

G is the Petersen graph, then $\Delta(G)=\gamma^*(G)=3$ while $\gamma(G)=4$. Seymour (1979) conjectured that for each, possibly nonsimple, graph one has $\gamma(G)\leqslant\max\{\Delta(G)+1, \lceil\gamma^*(G)\rceil\}$.

8. Cutting planes

For any set $P\subseteq\mathbb{R}^n$, let the *integer hull* of P, denoted by P_I, be

$$P_\mathrm{I}:=\operatorname{conv}\{x\mid x\in P, x \text{ integral}\}\,.$$

Trivially, if P is bounded, then P_I is a polytope. Meyer (1974) showed that if P is a rational polyhedron, then P_I is a rational polyhedron again.

Most of the combinatorial results given above consist of a characterization of the integer hull P_I by linear inequalities for certain polyhedra P. For example, the matching polytope is the integer hull of the polyhedron determined by the inequalities (5.28) (i), (ii). For most combinatorial optimization problems it is not difficult to describe a set of linear inequalities, determining a polyhedron P, in which the integral vectors are exactly the incidence vectors corresponding to the combinatorial optimization problem. Hence, P_I is the convex hull of these incidence vectors. However, it is generally difficult to describe P_I by linear inequalities (cf. section 9).

The *cutting-plane method* was introduced by Gomory (1960) to solve integer linear programs. Chvátal (1973a) (and Schrijver 1980a, for the unbounded case) derived from it the following iterative process characterizing P_I.

Define for any polyhedron $P\subseteq\mathbb{R}^n$:

$$P':=\bigcap_{\substack{H \text{ rational affine}\\ \text{halfspace with } H\supseteq P}} H_\mathrm{I}\,, \tag{8.2}$$

where a *rational affine halfspace* is a set $H:=\{x\mid c^\mathrm{T}x\leqslant\delta\}$, with $c\in\mathbb{Q}^n$ $(c\neq\mathbf{0})$ and $\delta\in\mathbb{Q}$. Clearly, we may assume that the components of c are relatively prime integers, which implies

$$H_\mathrm{I}=\{x\mid c^\mathrm{T}x\leqslant\lfloor\delta\rfloor\}\,. \tag{8.3}$$

This usually makes the set P' easy to characterize.

For instance, for any rational $m\times n$ matrix and $b\in\mathbb{Q}^m$ we have

$$\begin{aligned}&\{x\mid Ax\leqslant b\}'=\{x\mid(u^\mathrm{T}A)x\leqslant\lfloor u^\mathrm{T}b\rfloor \text{ for all } u\in\mathbb{Q}^m_+ \text{ with } u^\mathrm{T}A \text{ integral}\}\,,\\ &\{x\mid x\geqslant 0, Ax\leqslant b\}'=\{x\mid x\geqslant 0;\ \lfloor u^\mathrm{T}A\rfloor x\leqslant\lfloor u^\mathrm{T}b\rfloor \text{ for all } u\in\mathbb{Q}^m_+\}\end{aligned} \tag{8.4}$$

(here $\lfloor\cdot\rfloor$ denotes component-wise lower integer parts).

The halfspaces H_I (more strictly, their bounding hyperplanes) are called *cutting planes*.

It can be shown that if P is a rational polyhedron, then P' is also a rational polyhedron. Trivially, $P\subseteq H$ implies $P_\mathrm{I}\subseteq H_\mathrm{I}$, and hence $P_\mathrm{I}\subseteq P'$. Now generally $P''\neq P'$, and repeating this operation we obtain a sequence of polyhedra

$P, P', P'', P''', \ldots$, satisfying

$$P \supseteq P' \supseteq P'' \supseteq P''' \supseteq \cdots \supseteq P_{\mathrm{I}} . \tag{8.5}$$

Denote the $(t+1)$th set in this sequence by $P^{(t)}$. Then:

Theorem 8.6. *For each rational polyhedron P there exists a number t with $P^{(t)} = P_{\mathrm{I}}$.*

A direct consequence applies to bounded, but not necessarily rational, polyhedra.

Corollary 8.7. *For each polytope P there exists a number t with $P^{(t)} = P_{\mathrm{I}}$.*

Blair and Jeroslow (1982) (cf. Cook et al. 1986b) proved the following generalization of Theorem 8.6.

Theorem 8.8. *For each rational matrix A there exists a number t such that for each column vector b one has: $\{x \mid Ax \leqslant b\}^{(t)} = \{x \mid Ax \leqslant b\}_{\mathrm{I}}$.*

Hence we can define the *Chvátal rank* of a rational matrix A as the smallest such number t. The *strong Chvátal rank* of A then is the Chvátal rank of the matrix

$$\begin{bmatrix} I \\ -I \\ A \\ -A \end{bmatrix} . \tag{8.9}$$

It follows from Hoffman and Kruskal's Theorem (cf. Theorem 4.1) that an integral matrix A has a strong Chvátal rank 0 if and only if it is totally unimodular. Similar characterizations for higher Chvátal ranks are not known. In Examples 8.10 and 9.3 we shall see some classes of matrices with strong Chvátal rank 1.

For more on cutting planes, see Jeroslow (1978, 1979), and Blair and Jeroslow (1977, 1979, 1982).

Example 8.10 (*The matching polytope*). For any graph $G = (V, E)$, let P be the polytope determined by (5.28) (i), (ii). So P_{I} is the matching polytope of G. It is not difficult to show that P' is the polytope determined by (5.28) (i)–(iii). Hence Edmonds' Matching Polyhedron Theorem 5.33 is equivalent to asserting $P' = P_{\mathrm{I}}$. So the matching polytope arises from (5.28) (i), (ii) by one "round" of cutting planes.

It can be derived from Edmonds' Matching Polyhedron Theorem that each integer matrix $A = (a_{ij})$ satisfying

$$\sum_{i=1}^{m} |a_{ij}| \leq 2\,, \quad j = 1, \ldots, n\,, \tag{8.11}$$

where A has order $m \times n$, has strong Chvátal rank at most 1.

9. Hard problems and the complexity of the integer hull

The integer hull P_I can be quite intractable compared with the polyhedron P. This has been shown by Karp and Papadimitriou (1982), under the generally accepted assumption NP $\neq$ co-NP.

First note that the ellipsoid method (cf. section 3) can also be used in the negative: if NP $\neq$ P, then for any NP-complete problem there is no polynomial-time algorithm for the separation problem for the corresponding polytopes. More precisely, if for each graph $G = (V, E)$ $\mathscr{F}_G$ is a subset of $\mathscr{P}(E)$, and if Optimization Problem 3.17 is NP-complete, then (if NP $\neq$ P) the Separation Problem 3.18 is not polynomially solvable.

In fact, Karp and Papadimitriou showed that for any class $(\mathscr{F}_G \mid G$ graph), if Optimization Problem 3.17 is NP-complete, and if NP $\neq$ co-NP, then the class of polytopes $\mathrm{conv}\{\chi^F \mid F \in \mathscr{F}_G\}$ has difficult facets, i.e.,

> there exists no polynomial Φ such that for each graph $G = (V, E)$ and each $c \in \mathbb{Z}^E$ and $\delta \in \mathbb{Q}$ with the property that $c^{\mathrm{T}}x \leq \delta$ defines a facet of $\mathrm{conv}\{\chi^F \mid F \in \mathscr{F}_G\}$, the fact that $c^{\mathrm{T}}x \leq \delta$ is valid for each χ^F with $F \in \mathscr{F}_G$ has a proof of length at most $\Phi(|V| + |E| + \mathrm{size}(c) + \mathrm{size}(\delta))$. (9.1)

The meaning of (9.1) might become clear by considering description (5.28) of the matching polytope: although (5.28) consists of exponentially many inequalities, each facet-defining inequality is of form (5.28), and for them it is easy to show that they are valid for the matching polytope.

Another negative result was given by Boyd and Pulleyblank (1984): let, for a given class $(\mathscr{F}_G \mid G$ graph), for each graph $G = (V, E)$ the polytope P_G in $\mathbb{R}^E$ satisfy $(P_G)_I = \mathrm{conv}\{\chi^F \mid F \in \mathscr{F}_G\}$ and have the property that

$$\text{given } G = (V, E) \text{ and } c \in \mathbb{R}^E, \text{ find } \max\{c^{\mathrm{T}}x \mid x \in P_G\} \tag{9.2}$$

is polynomially solvable. Then if Optimization Problem 3.17 is NP-complete and if NP $\neq$ co-NP, then there is no fixed t such that for each graph G, $(P_G)^{(t)} = \mathrm{conv}\{\chi^F \mid F \in \mathscr{F}_G\}$.

Similar results hold for subcollections $\mathscr{F}_G$ of $\mathscr{P}(V)$ and for directed graphs. See also Papadimitriou (1984) and Papadimitriou and Yannakakis (1982) for the complexity of facets.

Example 9.3 (*The stable-set polytope*). Let $G = (V, E)$ be a graph, and let

STAB(G) be the stable-set polytope of G. Let $P(G)$ be the polytope in $\mathbb{R}^V$ determined by

$$\begin{aligned}&\text{(i)} \quad x_v \geq 0\,, \quad v \in V\,,\\ &\text{(ii)} \quad \sum_{v \in K} x_v \leq 1\,, \quad K \subseteq V,\ K \text{ clique}\,.\end{aligned} \tag{9.4}$$

So $P(G) = A(\mathrm{STAB}(\bar{G}))$ (cf. Section 7).

Clearly, $\mathrm{STAB}(G) \subseteq P(G)$. In fact, since the integral solutions to (9.4) are exactly the incidence vectors of stable sets, we have

$$\mathrm{STAB}(G) = P(G)_{\mathrm{I}}\,. \tag{9.5}$$

Chvátal (1973a, 1984) showed that there is no fixed t such that $P(G)^{(t)} = P(G)_{\mathrm{I}}$ for all graphs G (if NP ≠ co-NP, this follows from Boyd and Pulleyblank's result mentioned above), even if we restrict G to graphs with $\alpha(G) = 2$.

By Chvátal's Theorem 7.14, the class of graphs with $P(G)_{\mathrm{I}} = P(G)$ is exactly the class of perfect graphs. In Example 8.10 we mentioned Edmonds' result that if G is the line graph of some graph H, then $P(G)' = P(G)_{\mathrm{I}}$, which is the matching polytope of H.

The smallest t for which $P(G)^{(t)} = P(G)_{\mathrm{I}}$ is an indication of the computational complexity of the stability number $\alpha(G)$. Chvátal (1973) raised the question of whether there exists, for each fixed t, a polynomial-time algorithm determining $\alpha(G)$ for graphs G with $P(G)^{(t)} = P(G)_{\mathrm{I}}$. This is true for $t = 0$, i.e., for perfect graphs (Grötschel et al. 1981).

Minty (1980) and Sbihi (1978, 1980) extended Edmonds' result of the polynomial solvability of the maximum-weighted matching problem, by describing polynomial-time algorithms for finding a maximum-weighted stable set in $K_{1,3}$-free graphs (i.e., graphs with no $K_{1,3}$ as induced subgraph). Hence, by (3.9), the separation problem for stable-set polytopes of $K_{1,3}$-free graphs is polynomially solvable. Yet no explicit description of a linear inequality system defining STAB(G) for $K_{1,3}$-free graphs has been found. This would extend Edmonds' description of the matching polytope. It follows from Chvátal's result mentioned above that there is no fixed t such that $P(G)^{(t)} = P(G)_{\mathrm{I}}$ for all $K_{1,3}$-free graphs (see Giles and Trotter 1981).

Perhaps the most natural "relaxation" of the stable-set polytope of $G = (V, E)$ is the polytope $Q(G)$ determined by

$$\begin{aligned}&\text{(i)} \quad x_v \geq 0\,, \quad v \in V\,,\\ &\text{(ii)} \quad x_v + x_w \leq 1\,, \quad \{v, w\} \in E\,.\end{aligned} \tag{9.6}$$

Again, $Q(G)_{\mathrm{I}} = \mathrm{STAB}(G)$. Since $Q(G) \supseteq P(G)$, there is no t with $Q(G)^{(t)} = Q(G)_{\mathrm{I}}$ for all G. It is not difficult to see that $Q(G)'$ is the polytope determined by (9.6) together with

$$\sum_{v \in C} x_v \leq \frac{|C| - 1}{2}\,, \quad C \text{ is the vertex set of an odd circuit}\,. \tag{9.7}$$

It was shown by Gerards and Schrijver (1986) that if G has no subgraph H which arises from K_4 by replacing edges by paths such that each triangle in K_4 becomes an odd circuit in H, then $Q(G)' = \mathrm{STAB}(G)$. Graphs G with $Q(G)' = \mathrm{STAB}(G)$ are called by Chvátal (1975) *t-perfect*.

Gerards and Schrijver showed more generally the following. Let $A = (a_{ij})$ be an integral $m \times n$ matrix satisfying

$$\sum_{j=1}^{n} |a_{ij}| \leq 2\,, \quad i = 1, \ldots, m\,. \tag{9.8}$$

Then A has strong Chvátal rank at most 1 if and only if A cannot be transformed to the matrix

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \tag{9.9}$$

by a series of the following operations: deleting or permuting rows or columns, or multiplying them by -1; replacing $\left[\begin{smallmatrix} 1 & c^{\mathrm{T}} \\ b & D \end{smallmatrix}\right]$ by $D - bc^{\mathrm{T}}$, where D is a matrix and b and c are column vectors.

Chvátal (1973a) showed that for $G = K_n$ the smallest t with $Q(G)^{(t)} = \mathrm{STAB}(G)$ is about $\log n$.

Chvátal (1975) observed that the incidence vectors of two stable sets C, C' are adjacent on the stable-set polytope if and only if $C \,\Delta\, C'$ induces a connected graph. For more on the stable-set polytope, see Fulkerson (1971), Chvátal (1973a, 1975, 1984, 1985), Padberg (1973, 1974, 1977, 1979), Nemhauser and Trotter (1974, 1975), Trotter (1975), Wolsey (1976b), Balas and Zemel (1977), Ikura and Nemhauser (1985), Grötschel et al. (1986), and Lovász and Schrijver (1991).

Example 9.10 (*The traveling-salesman polytope*). For any graph $G = (V, E)$, the *traveling-salesman polytope* is equal to $\mathrm{conv}\{\chi^H \mid H \subseteq E,\ H$ Hamiltonian circuit$\}$. As the traveling salesman problem is NP-complete, by Karp and Papadimitriou's result, the traveling-salesman polytope will have "difficult" facets [cf. (9.1)] if NP $\neq$ co-NP.

Define the polyhedron $P \subseteq \mathbb{R}^E$ by:

$$\begin{aligned} &\text{(i)} && 0 \leq x_e \leq 1\,, \quad e \in E\,, \\ &\text{(ii)} && \sum_{e \ni v} x_e = 2\,, \quad v \in V\,, \\ &\text{(iii)} && \sum_{e \in \delta(U)} x_e \geq 2\,, \quad U \subseteq V,\ 3 \leq |U| \leq |V| - 3\,. \end{aligned} \tag{9.11}$$

Since the integral solutions of (9.11) are exactly the incidence vectors of Hamiltonian circuits, P_I is equal to the traveling-salesman polytope. Note that the problem of minimizing a linear function $c^{\mathrm{T}}x$ over P is polynomially solvable, with the ellipsoid method, since system (9.11) can be checked in polynomial time [(iii) can be checked by reduction to a minimum-cut problem]. So if NP$\neq$co-NP, by Boyd and Pulleyblank's result, there is no fixed t such that $P^{(t)} = P_{\mathrm{I}}$ for each graph G.

The system (9.11), however, has been useful in solving large-scale instances of the traveling salesman problem: for any $c \in \mathbb{Q}^E$, the minimum of $c^{\mathrm{T}}x$ over (9.11) is a lower bound for the traveling salesman problem, which can be computed with the simplex method using a row-generating technique. This lower bound can be used in a "branch-and-bound" procedure for the traveling salesman problem.

This approach was initiated by Dantzig et al. (1954, 1959), and developed and sharpened by Miliotis (1978), Grötschel and Padberg (1979a,b), Grötschel (1980), Crowder and Padberg (1980), and Padberg and Hong (1980) (see Grötschel and Padberg 1985, and Padberg and Grötschel 1985 for a survey).

Grötschel and Padberg (1979a) showed that the diameter of the traveling-salesman polytope for $G = K_n$ is equal to $\frac{1}{2}n(n-3)$. They also proved that for complete graphs all inequalities in (9.11) are facet-defining.

For more about facets of the traveling-salesman polytope, see Held and Karp (1970, 1971), Chvátal (1973b), Grötschel and Padberg (1975, 1977, 1979a,b), Maurras (1975), Grötschel (1977, 1980), Grötschel and Pulleyblank (1986), Grötschel and Wakabayashi (1981a,b), and Cornuéjols and Pulleyblank (1982).

Papadimitriou and Yannakakis (1982) showed that it is co-NP-complete to decide if a given vector belongs to the traveling-salesman polytope. Moreover, Papadimitriou (1978) showed that it is co-NP-complete to check if two Hamiltonian circuits H, H' yield adjacent incidence vectors (see also Rao 1976).

On the other hand, Padberg and Rao (1974) showed that the diameter of the "asymmetric" traveling-salesman polytope (i.e., convex hull of incidence vectors of Hamiltonian cycles in a directed graph) is equal to 2, for the complete directed graph with at least six vertices. Grötschel and Padberg (1985) conjecture that also the "undirected" traveling-salesman polytope has diameter 2.

Other hard problems 9.12. The following references deal with polyhedra associated with further difficult problems. *Set-packing problem*: Fulkerson (1971), Padberg (1973, 1977, 1979), Balas and Zemel (1977), and Ikura and Nemhauser (1985). *Set-covering problem*: Padberg (1979), Balas (1980), and Balas and Ho (1980). *Set-partitioning problem*: Balas and Padberg (1972), Balas (1977), Padberg (1979), and Johnson (1980). *Linear ordering and acyclic subgraph problem*: Grötschel et al. (1984, 1985a,b), and Jünger (1985). *Knapsack problem and* 0,1-*programming*: Balas (1975), Hammer et al. (1975), Wolsey (1975, 1976a, 1977), Johnson (1980), Zemel (1978), and Crowder et al. (1983). *Bipartite subgraph and maximum-cut problem*: Grötschel and Pulleyblank (1981), Barahona (1983a,b), and Barahona et al. (1985).

For more background information on hard problems, see Grötschel (1977, 1982).

References

Ahrens, J.H.
[1981] A conjecture of E.D. Bolker, *J. Combin. Theory B* **31**, 1–8.

Aráoz, J.
[1973] *Polyhedral neopolarities*, Ph.D. Thesis (University of Waterloo, Waterloo).

Aráoz, J., J. Edmonds and V.J. Griffin
[1983] Polarities given by systems of bilinear inequalities, *Math. Oper. Res.* **8**, 34–41.

Balas, E.
[1975] Facets of the knapsack polytope, *Math. Programming* **8**, 146 164.
[1977] Some valid inequalities for the set partitioning problem, *Ann. Discrete Math.* **1**, 13–47.
[1980] Cutting planes from conditional bounds: a new approach to set covering, *Math. Programming Stud.* **12**, 19–36.

Balas, E., and A. Ho
[1980] Set covering algorithms using cutting planes, heuristics and subgradient optimization: a computational survey, *Math. Programming Stud.* **12**, 37 60.

Balas, E., and M.W. Padberg
[1972] On the set-covering problem, *Oper. Res.* **20**, 1152–1161.

Balas, E., and E. Zemel
[1977] Critical cutsets of graphs and canonical facets of set-packing polytopes, *Math. Oper. Res.* **2**, 15 19.

Balinski, M.
[1983] Signatures des points extrêmes du polyèdre dual du problème de transport, *C.R. Acad. Sci. Paris Sér.* A-B **296**, 457–459.
[1984] The Hirsch conjecture for dual transportation polyhedra, *Math. Oper. Res.* 9, 629–633.

Balinski, M.L.
[1974] On two special classes of transportation polytopes, *Math. Programming Stud.* **1**, 43 58.
[1985] Signature methods for the assignment problem, *Oper. Res.* **33**, 527–536.

Balinski, M.L., and A. Russakoff
[1974] On the assigment polytope, *SIAM Rev.* **16**, 516–525.
[1984] Faces of dual transportation polyhedra, *Math. Programming Stud.* **22**, 1–8.

Barahona, F.
[1983a] The max-cut problem on graphs not contractible to K_5, *Oper. Res. Lett.* **2**, 107 111.
[1983b] On some weakly bipartite graphs, *Oper. Res. Lett.* **2**, 239–242.

Barahona, F., M. Grötschel and A.R. Mahjoub
[1985] Facets of the bipartite subgraph polytope, *Math. Oper. Res.* **10**, 340–358.

Baum, S., and L.E. Trotter Jr
[1977] Integer rounding and polyhedral decomposition of totally unimodular systems, in: *Optimization and Operations Research*, eds. R. Henn, B. Korte and W. Oettli (Springer, Berlin) pp. 15–23.

Bertsekas, D.P.
[1981] A new algorithm for the assignment problem, *Math. Programming* **21**, 152 171.

Birkhoff, G.
[1946] Tres observaciones sobre el algebra lineal, *Rev. Fac. Ci. Exactas, Puras y Aplicadas Univ. Nac. Tucuman, Ser. A* **5**, 147–151.

Blair, C.E., and R.G. Jeroslow
[1977] The value function of a mixed integer program: I, *Discrete Math.* **19**, 121–138.
[1979] The value function of a mixed integer program: II, *Discrete Math.* **25**, 7–19.
[1982] The value function of an integer program, *Math. Programming* **23**, 237 273.

Bland, R.G.
[1978] Elementary vectors and two polyhedral relaxations, *Math. Programming Stud.* **8**, 159–166.

Bolker, E.D.
[1972] Transportation polytopes, *J. Combin. Theory B* **13**, 251 262.

Boyd, S.C., and W.R. Pulleyblank

[1984] Facet generating techniques, to appear.

Camion, P.

[1965] Characterizations of totally unimodular matrices, *Proc. Amer. Math. Soc.* **16**, 1087–1073.

Chvátal, V.

[1973a] Edmonds polytopes and a hierarchy of combinatorial problems, *Discrete Math.* **4**, 305–337.

[1973b] Edmonds polytopes and weakly Hamiltonian graphs, *Math. Programming* **5**, 29–40.

[1975] On certain polytopes associated with graphs, *J. Combin. Theory B* **18**, 138–154.

[1984] *Cutting-plane Proofs and the Stability Number of a Graph*, Report 84326-OR (Institut für Operations Research, Universität Bonn, Bonn).

[1985] Cutting planes in combinatorics, *European J. Combin.* **6**, 217–226.

Cook, W.

[1983] Operations that preserve total dual integrality, *Oper. Res. Lett.* **2**, 31–35.

[1986] On box totally dual integral polyhedra, *Math. Programming* **34**, 48–61.

Cook, W., L. Lovász and A. Schrijver

[1984] A polynomial-time test for total dual integrality in fixed dimension, *Math. Programming Stud.* **22**, 64–69.

Cook, W., J. Fonlupt and A. Schrijver

[1986a] An integer analogue of Carathéodory's theorem, *J. Combin. Theory B* **40**, 63–70.

Cook, W., A.M.H. Gerards, A. Schrijver and É. Tardos

[1986b] Sensitivity theorems in integer linear programming, *Math. Programming* **34**, 251–264.

Cornuéjols, G., and W.R. Pulleyblank

[1982] The travelling salesman polytope and $\{0,2\}$-matching, *Ann. Discrete Math.* **16**, 27–55.

Crowder, H., and M.W. Padberg

[1980] Solving large-scale symmetric travelling salesman problems to optimality, *Management Sci.* **26**, 495–509.

Crowder, H., E.L. Johnson and M.W. Padberg

[1983] Solving large-scale zero–one linear programming problems, *Operations Res.* **31**, 803–834.

Cunningham, W.H.

[1979] Theoretical properties of the network simplex method, *Math. Oper. Res.* **4**, 196–208.

[1984] Testing membership in matroid polyhedra, *J. Combin. Theory B* **36**, 161–188.

Cunningham, W.H., and A.B. Marsh III

[1978] A primal algorithm for optimal matching, in: *Polyhedral Combinatorics (dedicated to the memory of D.R. Fulkerson)*, eds. M.L. Balinski and A.J. Hoffman, *Math. Programming Stud.* **8**, 50–72.

Dantzig, G.B.

[1951a] Maximization of a linear function of variables subject to linear inequalitities, in: *Activity Analysis of Production and Allocation*, ed. Tj.C. Koopmans (Wiley, New York) pp. 339–347.

[1951b] Application of the simplex method to a transportation problem, in: *Activity Analysis of Production and Allocation*, ed. Tj.C. Koopmans (Wiley, New York) pp. 359–373.

[1963] *Linear Programming and Extensions* (Princeton University Press, Princeton, NJ).

Dantzig, G.B., D.R. Fulkerson and S.M. Johnson

[1954] Solution of a large-scale traveling-salesman problem, *Oper. Res.* **2**, 393–410.

Dantzig, G.B., L.R. Ford and D.R. Fulkerson

[1956] A primal-dual algorithm for linear programs, in: *Linear Inequalities and Related Systems*, eds. H.W. Kuhn and A.W. Tucker (Princeton University Press, Princeton, NJ) pp. 171–181.

Dantzig, G.B., D.R. Fulkerson and S.M. Johnson

[1959] On a linear programming, combinatorial approach to the travelling-salesman problem, *Oper. Res.* **7**, 58–66.

Dinits, E.A.

[1970] Algorithm for solution of a problem of maximum flow in a network with power estimation, *Soviet Math. Dokl.* **11** (1977) 1277–1280.

Edmonds, J.

[1965] Maximum matching and a polyhedron with 0, 1-vertices, *J. Res. Nat. Bur. Standards B* **69**, 125–130.

[1967] Optimum branchings, *J. Res. Nat. Bur. Standards B* **71**, 233–240.

[1970] Submodular functions, matroids and certain polyhedra, in: *Combinatorial Structures and Their Applications*, ed. R. Guy (Gordon and Breach, New York) pp. 69–87.

[1971] Matroids and the greedy algorithm, *Math. Programming* **1**, 127–136.

[1973] Edge-disjoint branchings, in: *Combinatorial Algorithms*, ed. R. Rustin (Academic Press, New York) pp. 91–96.

[1979] Matroid intersection, *Ann. Discrete Math.* **4**, 39–49.

Edmonds, J., and R. Giles

[1977] A min–max relation for submodular functions on graphs, *Ann. Discrete Math.* **1**, 185–204.

[1984] Total dual integrality of linear inequality systems, in: *Progress in Combinatorial Optimization*, ed. W.R. Pulleyblank (Academic Press, Toronto) pp. 117–129.

Edmonds, J., and R.M. Karp

[1972] Theoretical improvements in algorithmic efficiency for network flow problems, *J. Assoc. Comput. Mach.* **19**, 248–264.

Egerváry, J.

[1931] Matrixok kombinatorius tuladonságairól, *Mat. Fiz. Lapok* **38**, 16–28.

Ford Jr, L.R., and D.R. Fulkerson

[1957] A simple algorithm for finding maximal network flows and an application to the Hitchcock problem, *Canad. J. Math.* 9, 210–218.

Frank, A.

[1980] On the orientation of graphs, *J. Combin. Theory B* **28**, 251–261.

Frank, A., and É. Tardos

[1984] Matroids from crossing families, in: *Finite and Infinite Sets I*, eds. A. Hajnal, R. Rado and V.T. Sós (North-Holland, Amsterdam) pp. 295–304.

[1985] An application of simultaneous approximation in combinatorial optimization, in: *26th Annu. Symp. on Foundations of Computer Science* (IEEE Computer Society Press, New York) pp. 459–463.

Fulkerson, D.R.

[1970a] Blocking polyhedra, in: *Graph Theory and its Applications*, ed. B. Harris (Academic Press, New York) pp. 93–112.

[1970b] The perfect graph conjecture and pluperfect graph theorem, in: *Proc. 2nd Chapel Hill Conf. on Combinatorial Mathematics and Its Applications*, eds. R.C. Bose et al. (University of North Carolina, Chapel Hill, NC) pp. 171–175.

[1971] Blocking anti-blocking pairs of polyhedra, *Math. Programming* **1**, 168–194.

[1972] Anti-blocking polyhedra, *J. Combin. Theory B* **12**, 50–71.

[1973] On the perfect graph theorem, in: *Mathematical Programming*, eds. T.C. Hu and S.M. Robinson (Academic Press, New York) pp. 69–76.

[1974] Packing rooted directed cuts in a weighted directed graph, *Math. Programming* **6**, 1–13.

Gerards, A.M.H., and A. Schrijver

[1986] Matrices with the Edmonds–Johnson property, *Combinatorica* **6**, 365–379.

Ghouila-Houri, A.

[1962] Caractérisation des matrices totalement unimodulaires, *C.R. Acad. Sci. Paris* **254**, 1192–1194.

Giles, F.R.

[1975] *Submodular functions, graphs and integer polyhedra*, Ph.D. Thesis (University of Waterloo, Waterloo, Ont.).

Giles, F.R., and W.R. Pulleyblank

[1979] Total dual integrality and integral polyhedra, *Linear Algebra Appl.* **25**, 191–196.

Giles, R.

[1978] Facets and other faces of branching polyhedra, in: *Combinatorics I*, eds. A. Hajnal and V.T. Sós (North-Holland, Amsterdam) pp. 401–418.

Giles, R., and L.E. Trotter Jr

[1981] On stable set polyhedra for $K_{1,3}$-free graphs, *J. Combin. Theory B* **31**, 313–326.

Goldfarb, D.

[1985] Efficient dual simplex algorithms for the assignment problem, *Math. Programming* **33**, 187–203.

Gomory, R.E.

[1960] Solving linear programs in integers, in: *Combinatorial Analysis*, eds. R. Bellman and M. Hall Jr (American Mathematical Society, Providence, RI) pp. 211–215.

[1963] An algorithm for integer solutions to linear programs, in: *Recent Advances in Mathematical Programming*, eds. R.L. Graves and P. Wolfe (McGraw-Hill, New York) pp. 269–302.

Griffin, V.

[1977] *Polyhedral polarity*, Ph.D. Thesis (University of Waterloo, Waterloo, Ont.).

Griffin, V., J. Aráoz and J. Edmonds

[1982] Polyhedral polarity defined by a general bilinear inequality, *Math. Programming* **23**, 117–137.

Grötschel, M.

[1977] *Polyedrische Charakterisierungen kombinatorischer Optimierungsprobleme* (Verlag Anton Hain, Meisenheim am Glan).

[1980] On the symmetric travelling salesman problem: solution of a 120-city problem, *Math. Programming Stud.* **12**, 61–77.

[1982] Approaches to hard combinatorial optimization problems, in: *Modern Applied Mathematics Optimization and Operations Research*, ed. B. Korte (North-Holland, Amsterdam) pp. 437–515.

Grötschel, M., and M.W. Padberg

[1975] Partial linear characterization of the asymmetric travelling salesman polytope, *Math. Programming* **8**, 378–381.

[1977] Lineare Charakterisierungen von Travelling Salesman Problemen, *Z. Oper. Res.* **21**, 33–64.

[1979a] On the symmetric travelling salesman problem I: Inequalities, *Math. Programming* **16**, 265–280.

[1979b] On the symmetric travelling salesman problem I: Lifting theorems and facets, *Math. Programming* **16**, 281–302.

[1985] Polyhedral theory, in: *The Traveling Salesman Problem, A Guided Tour through Combinatorial Optimization*, eds. E.L. Lawler, J.K. Lenstra, A.H.G. Rinnooy Kan and D.B. Shmoys (Wiley, Chichester) pp. 251–305.

Grötschel, M., and W.R. Pulleyblank

[1981] Weakly bipartite graphs and the max-cut problem, *Oper. Res. Lett.* **1**, 23–27.

[1986] Clique tree inequalities and the symmetric travelling salesman problem, *Math. Oper. Res.* **11**, 537–569.

Grötschel, M., and Y. Wakabayashi

[1981a] On the structure of the monotone asymmetric travelling salesman polytope I: hypohamiltonian facets, *Discrete Math.* **34**, 43–59.

[1981b] On the structure of the monotone asymmetric travelling salesman polytope II: hypotraceable facets, *Math. Programming Stud.* **14**, 77–97.

Grötschel, M., L. Lovász and A. Schrijver

[1981] The ellipsoid method and its consequences in combinatorial optimization, *Combinatorica* **1**, 169–197.

Grötschel, M., M. Jünger and G. Reinelt

[1984] A cutting plane algorithm for the linear ordering problem, *Oper. Res.* **32**, 1195–1220.

[1985a] On the acyclic subgraph polytope, *Math. Programming* **33**, 1–27.

[1985b] Facets of the linear ordering polytope, *Math. Programming* **33**, 43–60.

Grötschel, M., L. Lovász and A. Schrijver

[1986] Relaxations of vertex packing, *J. Combin. Theory B* **40**, 330–343.

[1988] *Geometric Algorithms and Combinatorial Optimization* (Springer, Heidelberg).

Grünbaum, B.

[1967] *Convex Polytopes* (Interscience-Wiley, London).

Gupta, R.P.

[1967] A decomposition theorem for bipartite graphs, in: *Theory of Graphs*, ed. P. Rosenstiehl (Gordon and Breach, New York) pp. 135–138.

Hammer, P.L., E.L. Johnson and U.N. Peled

[1975] Facets of regular 0–1 polytopes, *Math. Programming* **8**, 179–206.

Hausmann, D., and B. Korte

[1978] Colouring criteria for adjacency on 0–1 polyhedra, *Math. Programming Stud.* **8**, 106–127.

Held, M., and R.M. Karp

[1970] The traveling-salesman problem and minimum spanning trees, *Oper. Res.* **18**, 1138–1162.

[1971] The traveling-salesman problem and minimum spanning trees: Part II, *Math. Programming* **1**, 6–25.

Hitchcock, F.L.

[1941] The distribution of a product from several sources to numerous localities, *J. Math. Phys.* **20**, 224–230.

Hoffman, A.J.

[1960] Some recent applications of the theory of linear inequalities to extremal combinatorial analysis, in: *Combinatorial Analysis*, eds. R. Bellman and M. Hall Jr (American Mathematical Society, Providence, RI) pp. 113–127.

[1974] A generalization of max flow–min cut, *Math. Programming* 6, 352–359.

[1979] The role of unimodularity in applying linear inequaltities to combinatorial theorems, *Ann. Discrete Math.* **4**, 73–84

Hoffman, A.J., and J.B. Kruskal

[1956] Integral boundary points of convex polyhedra, in: *Linear Inequalities and Related Systems*, eds. H.W. Kuhn and A.W. Tucker (Princeton University Press, Princeton, NJ) pp. 223–246.

Huang, H.-C., and L.E. Trotter Jr

[1980] A technique for determining blocking and anti-blocking polyhedral descriptions, *Math. Programming Stud.* **12**, 197–205.

Hung, M.S.

[1983] A polynomial simplex method for the assignment problem, *Oper. Res.* **31**, 595–600.

Hurkens, C.A.J.

[1991] On the diameter of the edge cover polytope, *J. Combin. Theory B* **51**, 271–276.

Ikura, Y., and G.L. Nemhauser

[1983] *A Polynomial-time Dual Simplex Algorithm for the Transportation Problem*, Tech. Report 602 (School of Operations Research and Information Engineering, Cornell University, Ithaca, NY).

[1985] Simplex pivots on the set packing polytope, *Math. Programming* **33**, 123–138.

Jeroslow, R.

[1979] An introduction to the theory of cutting-planes, *Ann. Discrete Math.* **5**, 71–95.

Jeroslow, R.G.

[1978] Cutting-plane theory: algebraic methods, *Discrete Math.* **23**, 121–150.

Johnson, E.L.

[1978] Support functions, blocking pairs and anti-blocking pairs, *Math. Programming Stud.* **8**, 167–196.

[1980] Subadditive lifting methods for partitioning and knapsack problems, *J. Algorithms* **1**, 75–96.

Jünger, M.

[1985] *Polyhedral Combinatorics and the Acyclic Subgraph Problem* (Heldermann, Berlin).

Karmarkar, N.

[1984] A new polynomial-time algorithm for linear programming, *Combinatorica* **4**, 373–395.

Karp, R.M., and C.H. Papadimitriou

[1982] On linear characterizations of combinatorial optimization problems, *SIAM J. Comput.* **11**, 620–632.

Khachiyan, L.G.

[1979] A polynomial algorithm in linear programming, *Souiet Math. Dokl.* **20**, 191–194.

Klee, V., and D.W. Walkup

[1967] The d-step conjecture for polyhedra of dimension $d < 6$, *Acta Math. (Uppsala)* **117**, 53–78.

Klee, V., and C. Witzgall

[1968] Facets and vertices of transportation polyhedra, in: *Mathematics of the Decision Sciences, Part I*, eds. G.B. Dantzig and A.F. Veinott (American Mathematical Society, Providence, RI) pp. 257–282.

Koopmans, Tj.C.

[1948] Optimum utilization of the transportation system, in: *The Econometric Society Meeting, Washington, DC, 1947*, ed. D.H. Leavens, pp. 136–146.

Larman, D.G.

[1970] Paths on polytopes, *Proc. London Math. Soc. (3)* **20**, 161–178.

Lehman, A.

[1965] On the width-length inequality, Mimeographic notes. See: 1979, *Math. Programming* 7, 403–417.

Lenstra, A.K., H.W. Lenstra Jr and L. Lovász

[1982] Factoring polynomials with rational coefficients, *Math. Ann.* **261**, 515–534.

Lovász, L.

[1972] Normal hypergraphs and the perfect graph conjecture, *Discrete Math.* **2**, 253–267.

[1976] On two minimax theorems in graphs, *J. Combin. Theory B* **21**, 96–103.

[1977] Certain duality principles in integer programming, *Ann. Discrete Math.* **1**, 363–374.

[1979] Graph theory and integer programming, *Ann. Discrete Math.* **4**, 141–158.

Lovász, L., and M.D. Plummer

[1986] *Matching Theory, Ann. Discrete Math.* **29**.

Lovász, L., and A. Schrijver

[1991] Cones of matrices and setfunctions and 0, 1 optimization, *SIAM J. Optim.* **1**, 166–190.

Lucchesi, C.L., and D.H. Younger

[1978] A minimax relation for directed graphs, *J. London Math. Soc. (2)* **17**, 369–374.

Maurras, J.F.

[1975] Some results on the convex hull of Hamiltonian cycles of symmetric complete graphs, in: *Combinatorial Programming: Methods and Applications* (Reidel, Dordrecht) pp. 179–190.

Meyer, R.R.

[1974] On the existence of optimal solutions to integer and mixed-integer programming problems, *Math. Programming* **7**, 223–235.

Miliotis, P.

[1978] Using cutting planes to solve the symmetric travelling salesman problem, *Math. Programming* **15**, 177–188.

Minkowski, H.

[1896] *Geometrie der Zahlen,* Erste Lieferung (Teubner, Leipzig).

Minty, G.J.

[1980] On maximal independent sets of vertices in claw-free graphs, *J. Combin. Theory B* **28**, 284–304.

Motzkin, T.S.

[1936] *Beiträge zur Theorie der linearen Ungleichungen,* Inaugural Dissertation, Basel (Azriel, Jerusalem).

Naddef, D.

[1989] The Hirsch conjecture is true for (0, 1)-polytopes, *Math. Programming* **45**, 109–110.

Nemhauser, G.L., and L.E. Trotter Jr

[1974] Properties of vertex packing and independence system polyhedra, *Math. Programming* **6**, 48–61.

[1975] Vertex packings: structural properties and algorithms, *Math. Programming* **8**, 232–248.

Padberg, M.W.

[1973] On the facial structure of set packing polyhedra, *Math. Programming* **5**, 199–215.

[1974] Perfect zero–one matrices, *Math. Programming* **6**, 180–196.

[1977] On the complexity of set packing polyhedra, *Ann. Discrete Math.* **1**, 421–434.

[1979] Covering, packing and knapsack problems, *Ann. Discrete Math.* **4**, 265–287.

Padberg, M.W., and M. Grötschel

[1985] Polyhedral computations, in: *The Traveling Salesman Problem, A Guided Tour through Combinatorial Optimization,* eds. E.L. Lawler, J.K. Lenstra, A.H.G. Rinnooy Kan and D.B. Shmoys (Wiley, Chichester) pp. 307–360.

Padberg, M.W., and S. Hong

[1980] On the symmetric travelling salesman problem: a computational study, *Math. Programming Stud.* **12**, 78–107.

Padberg, M.W., and M.R. Rao

[1974] The travelling salesman problem and a class of polyhedra of diameter two, *Math. Programming* **7**, 32–45.

[1980] *The Russian Method and Integer Programming*, GBA Working Paper (New York University, New York).

[1982] Odd minimum cut-sets and *b*-matchings, *Math. Oper. Res.* 7, 67–80.

Papadimitriou, C.H.

[1978] The adjacency relation on the traveling salesman polytope is NP-complete, *Math. Programming* **14**, 312–324.

[1984] Polytopes and complexity, in: *Progress in Combinatorial Optimization*, ed. W.R. Pulleyblank (Academic Press, Toronto) pp. 295–305.

Papadimitriou, C.H., and K. Steiglitz

[1982] *Combinatorial Optimization: Algorithms and Complexity* (Prentice-Hall, Englewood Cliffs, NJ).

Papadimitriou, C.H., and M. Yannakakis

[1982] The complexity of facets (and some facets of complexity), in: *Proc. 14th Annu. ACM Symp. on Theory of Computing* (ACM, New York) pp. 255–260.

Pulleyblank, W.R.

[1983] Polyhedral combinatorics, in: *Mathematical Programming – The State of the Art*, eds. A. Bachem, M. Grötschel and B. Korte (Springer, Berlin) pp. 312–345.

Rao, M.R.

[1976] Adjacency of the traveling salesman tours and 0–1 vertices, *SIAM J. Appl. Math.* **30**, 191–198.

Roohy-Laleh, E.

[1981] *Improvements to the theoretical efficiency of the network simplex method*, Ph.D. Thesis (Carleton University, Ottawa).

Saigal, R.

[1969] A proof of the Hirsch conjecture on the polyhedron of the shortest route problem, *SIAM J. Appl. Math.* **17**, 1232–1238.

Sbihi, N.

[1978] *Étude des stables dans les graphes sans étoile*, M.Sc. Thesis (Université Sci. et Méd. de Grenoble, Grenoble).

[1980] Algorithme de recherche d'un stable de cardinalité maximum dans un graphe sans étoile, *Discrete Math.* **29**, 53–76.

Schrijver, A.

[1980a] On cutting planes, *Ann. Discrete Math.* **9**, 291–296.

[1980b] A counterexample to a conjecture of Edmonds and Giles, *Discrete Math.* **33**, 213–214.

[1981] On total dual integrality, *Linear Algebra Appl.* **38**, 27–32.

[1982] Min–max relations for directed graphs, *Ann. Discrete Math.* **16**, 261–280.

[1983a] Packing and covering of crossing families of cuts, *J. Combin. Theory B* **35**, 104–128.

[1983b] Min–max results in combinatorial optimization, in: *Mathematical Programming – The State of the Art*, eds. A. Bachem, M. Grötschel and B. Korte (Springer, Berlin) pp. 439–500.

[1986] *Theory of Linear and Integer Programming* (Wiley, Chichester).

Seymour, P.D.

[1977] The matroids with the max-flow min-cut property, *J. Combin. Theory B* **23**, 189–222.

[1979] On multicolourings of cubic graphs and conjectures of Fulkerson and Tutte, *Proc. London Math. Soc. (3)* **38**, 423–460.

[1980] Decomposition of regular matroids, *J. Combin. Theory B* **28**, 305–359.

Shor, N.Z.

[1970a] Utilization of the operation of space dilatation in the minimization of convex functions, *Cybernetics* **6**, 7–15.

[1970b] Convergence rate of the gradient decent method with dilatation of the space, *Cybernetics* **6**, 102–108.

[1977] Cut-off method with space extension in convex programming problems, *Cybernetics* **13**, 94–96.

Steinitz, E.

[1916] Bedingt konvergente Reihen und konvexe Systeme (Schluss), *J. Reine Angew. Math.* **146**, 1–52.

Stoer, J., and C. Witzgall

[1970] *Convexity and Optimization in Finite Dimensions I* (Springer, Berlin).

Trotter Jr, L.E.
[1975] A class of facet producing graphs for vertex packing polyhedra, *Discrete Math.* **12**, 373–388.

Truemper, K.
[1982] On the efficiency of representability tests for matroids, *European J. Combin.* **3**, 275–291.
[1990] A decomposition theorem for matroids. V. Testing of matrix total unimodularity, *J. Combin. Theory B* 49, 241–281.

Weyl, H.
[1935] Elementare Theorie der konvexen Polyeder, *Comm. Math. Helv.* **7**, 290–306.

Wolsey, L.A.
[1975] Faces for a linear inequality in 0–1 variables, *Math. Programming* **8**, 165–178.
[1976a] Facets and strong valid inequalities for integer programs, *Oper. Res.* **24**, 367–372.
[1976b] Further facet generating procedures for vortex packing polytopes, *Math. Programming* **11**, 158–163.
[1977] Valid inequalities and superadditivity for 0–1 integer programs, *Math. Oper. Res.* **2**, 66–77.

Yudin, D.B., and A.S. Nemirovskiĭ
[1976/1977] Evaluation of the informational complexity of mathematical programming problems, *Matekon* **13**(2), 3–25.
[1977] Informational complexity and efficient methods for the solution of convex extremal problems, *Matekon* **13**(3), 25–45.

Zemel, E.
[1978] Lifting the facets of zero–one polytopes, *Math. Programming* **15**, 268–277.

Zhu, Y.-J.
[1963] Maximum number of iterations in the dual algorithm of the Kantorovic–Hitchcock problem in linear programming, *Chinese Math.* **3**, 307–313.

CHAPTER 31

Tools from Linear Algebra

C.D. GODSIL

Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ont. N2L 3G1, Canada

with an Appendix by

L. LOVÁSZ

Department of Computer Science, Yale University, New Haven, CT 06250, USA

Contents

HANDBOOK OF COMBINATORICS
Edited by R. Graham, M. Grötschel and L. Lovász

1. Introduction

Linear algebra provides an important collection of tools for the working combinatorialist. These have often been used to obtain the first, the most elegant, or the only proof of many significant results. Before I compiled this survey, my opinion was that this area consisted of a large collection of disparate "tricks". I have since come around to the view that there is a small set of basic principles, perhaps not easily formalised, that underly most of the combinatorial applications of linear algebra.

In writing this survey I have made no attempt to be exhaustive; indeed I should apologise in advance to each of my readers for leaving out their favourite example. The references provided are also far from complete, but should at least form a reasonable starting point for those wishing to learn more.

The reader is hereby warned that, unless explicitly mentioned otherwise, all ranks, dimensions etc. are over the rationals. The letters I and J will denote the identity matrix and the "all-ones" matrix respectively. Their order will be determined by the context.

2. The rank argument

The best known application of linear algebra to combinatorics is the now standard proof of Fisher's inequality, namely that in any non-trivial 2-design the number of blocks is at least as large as the number of points. This seems a good place for us to begin. We first need to set up some notation. A *hypergraph* $H = (V, E)$ consists of a vertex set V and a collection E of subsets of V, which we call edges. We call H *simple* if there are no repeated edges and we say it is *k-uniform*, or just uniform, if each edge contains exactly k vertices. If each vertex of H lies in exactly r edges then H is *r-regular*, or simply regular. A simple 2-uniform hypergraph is better known as a "graph".

A t-design is a uniform hypergraph with the property that every subset of t vertices is contained in exactly λ edges, for some constant λ. Thus a 1-design is a λ-regular uniform hypergraph. It is well known and simple to prove that any t-design is also an s-design, for all s less than or equal to t. A design is *trivial* if each edge contains all the vertices. For further background see A. Brouwer's chapter on designs in this Handbook. Fisher's inequality (Fisher 1940) asserts, in our notation, that every non-trivial 2-design has at least as many edges as vertices. To prove this using linear algebra requires the use of incidence matrices, and consequently another definition.

The *incidence matrix* $B = B(H)$ of a hypergraph is the 01-matrix with rows indexed by the vertices of H, columns indexed by the edges, and with $(B)_{ij} = 1$ if and only if vertex i is contained in edge j. The rank of B can not be greater than either the number of rows or the number of columns of B. Thus we have the following.

Principle 2.1. *Let $H = (V, E)$ be a hypergraph with incidence matrix B. If the rows of B are linearly independent then $|V| \leqslant |E|$.*

This result is simultaneously too important, and too useful, to be termed a theorem. There is one problem remaining though: it is still up to us to determine the rank of the incidence matrix B. For an arbitrary large hypergraph this would normally be every bit as difficult as proving that $|V| \leqslant |E|$ by any other means. What saves us is that, in many interesting cases, the rank of $B(H)$ is more or less obvious due, for example, to some regularity in the structure of H. Thus in the case of 2-designs we find that the defining conditions imply that

$$BB^{\mathrm{T}} = (r - \lambda)I + \lambda J, \tag{2.1}$$

where r and λ are as above. If the block size k of the design is not equal to $|V|$ then we must have $r > \lambda$. Hence the right-hand side of (2.1) is the sum of a positive semi-definite matrix J and a positive definite matrix $(r - \lambda)I$. It is therefore positive-definite and, with that, non-singular. Consequently the left-hand side of (2.1) is non-singular, implying that the rank of B is equal to the number of rows in B. This proves Fisher's inequality. (Note that the use of positive-definiteness in the above argument can be circumvented by explicitly computing the determinant of $(r - \lambda)I + \lambda J$.)

We note further that, if B has rank $|V|$ then it contains a $|V| \times |V|$ submatrix with non-zero determinant. Given the definition of the determinant as a signed sum of products of entries of a matrix, we deduce that there is an injection $\phi : V \to E$ such that the edge $\phi(i)$ contains the vertex i, for all vertices i in H. This is a strengthening of the bald statement that $v \leqslant e$. If we replace the non-zero elements of B by distinct members from a set of algebraically independent numbers, we obtain a "generic" incidence matrix for H. The existence of a bijection of the type described is equivalent to requiring that the rank of this generic incidence matrix be equal to $|V|$. (For another, important, example of this type of argument, see Stanley 1980.)

Fisher's inequality can be generalised in many ways. If we weaken our definition of 2-design by allowing the edges to contain differing numbers of vertices, we find that B satisfies the matrix equation

$$BB^{\mathrm{T}} = \Delta + \lambda J, \tag{2.2}$$

where Δ is a diagonal matrix with non-negative entries. The diagonal entries of Δ will be positive if, for each pair of vertices in H, there is an edge containing one but not the other. In this case the argument we used above still yields that B has rank equal to $|V|$, and hence that $v \leqslant e$. (This result is due to Majindar 1962, and De Caen and Gregory 1985 prove an even more general result using quadratic forms.)

Another important generalisation of Fisher's inequality arises if we introduce automorphism groups. Suppose that Γ is a group of automorphisms of our hypergraph H. Then vertices and edges of H are partitioned into orbits by Γ. If H is a 2-design or, more generally, if B has rank $|V|$, then the number of edge orbits of Γ is as least as large as the number of vertex orbits. (If Γ is the identity group then this is just Fisher's inequality again.) This claim can be proved as follows. Let $C_1, \dots, C_k$ denote the vertex orbits of Γ. Call two edges σ and τ *equivalent* if $|\sigma \cap C_i| = |\tau \cap C_i|$ for all i. Clearly any two edges in the same edge orbit of Γ are equivalent. Let P be the $k \times v$ matrix with ith row equal to the characteristic vector of C_i (viewed as a subset of $V(H)$). Then edges σ and τ are equivalent if and only if the corresponding columns of PB are equal. Hence the number of edge orbits of Γ is as least as large as the rank of PB. If B has rank $|V|$ then $x^{\mathrm{T}}PB = 0$ if and only if $x^{\mathrm{T}}P = 0$. As the rows of P are linearly independent it follows that $x^{\mathrm{T}}P = 0$ if and only if $x = 0$, i.e., the rank of PB is equal to the number of rows of P. This proves our claim.

The argument used in the last paragraph is sufficiently important to be worth formalising. Let H be an arbitrary hypergraph, let π be a partition of its vertex set and let ρ be a partition of its edge set. Define the *characteristic matrix* of a partition to be the matrix with the ith row equal to the characteristic vector of the ith cell (or component) of the partition. (Thus, a 01-matrix is the characteristic matrix of a partition of its columns if and only if the sum of its rows is the vector with all entries equal to 1.) Denote the characteristic matrices of π and ρ by P and R respectively. We call the pair (π, ρ) of partitions equitable if:

(a) each edge in the jth cell of ρ contains the same number of vertices from the ith cell of π,

(b) each vertex in the ith cell of π is contained in the same number of edges from the jth cell of ρ.

We see that (π, ρ) is an equitable partition of H if and only if (ρ, π) is an equitable partition of the dual hypergraph. (This is obtained by swapping the roles of the vertices and edges in H – its incidence matrix is the transpose of that of H.)

Lemma 2.2. *Let π and ρ respectively be partitions of the vertices and edges of the hypergraph H. Then (π, ρ) is equitable if and only if there are matrices Φ and Ψ such that $PB = \Phi R$ and $RB^{\mathrm{T}} = \Psi P$.*

Proof. This lemma is only a routine translation of the definition (into linear algebra terms). □

If Φ and Ψ exist as described then $\Phi RR^{\mathrm{T}} = PBR^{\mathrm{T}}$ and $\Psi PP^{\mathrm{T}} = RB^{\mathrm{T}}P^{\mathrm{T}}$. Hence

$$\Phi RR^{\mathrm{T}} = PP^{\mathrm{T}}\Psi^{\mathrm{T}}.$$

Thus Ψ is determined by Φ, and vice versa. Note that both PP^{T} and RR^{T} are diagonal matrices. We call the matrix Φ the *vertex quotient* of B with respect to the given pair of partitions.

Lemma 2.3. *Let Φ be a vertex quotient of the incidence matrix B with respect to the equitable pair of partitions (π, ρ). If the rows of B are linearly independent then the rank of Φ is equal to the number of cells in π, and so the number of cells of π is less than or equal to the number of cells of ρ.*

Proof. We have

$$\operatorname{rank}(P) = \operatorname{rank}(PB) = \operatorname{rank}(\Phi R) = \operatorname{rank}(\Phi),$$

where the first and third equalities hold because the rows of P and R are linearly independent, while the second equality follows from Lemma 2.2. □

Note that Lemma 2.3 is actually a generalisation of Principle 2.1, which we can recover by taking π and ρ to be the partitions with all cells singletons. One important consequence of this lemma is the fact that the number of point orbits of a collineation group of a projective plane is always less than or equal to the number of line orbits (Hughes and Piper 1973, Theorem 13.4). It is not difficult to extend Lemma 2.3 to infinite structures. (See Cameron 1976.) The notion of quotient is useful because it provides a means of arguing that a particular matrix Φ has rank equal to the number of rows in it. (Namely, Φ has inherited this property from the larger matrix B.) Thus quotients extend the applicability of the rank argument. They will also play an important role in our section on eigenvalue methods. The definitions above have been chosen with this later usage in mind as well.

I should also mention that it is often convenient to view Φ as a generalised incidence matrix for the "quotient hypergraph" with the cells of π and ρ of H as its vertices and edges. (A cell of π is incident with a cell of ρ whenever some vertex in the former is contained in some edge of the latter.) In the case when π and ρ are the vertex and edge orbits of a group of automorphisms of H, Lemma 2.3 is well known and can be stated in a sharper form. See, e.g., Dembowski (1968, p. 22) and Stanley (1982, Lemma 9.1).

The next result is of fundamental importance, and underlies many combinatorial applications of linear algebra.

Theorem 2.4. *Let Ω be a set with cardinality n and let B be the incidence matrix for the hypergraph H with the k-sets of Ω as its vertices and the l-sets as its edges. Then if $k \leqslant \min\{l, n-l\}$, the rows of B are linearly independent.*

Here a k-set is incident with an l-set if it is contained in it. The earliest proof of this known to the writer appears in Gottlieb (1966). Other proofs appear in Foody and Hedayat (1977), Kantor (1972) and Graham et al. (1980). It can also be derived by a quotient argument. For suppose that we have a non-zero vector x such that $x^{\mathrm{T}}B = 0$. We may assume without loss that the first entry of x is non-zero; in fact we assume that it is equal to 1. Clearly Sym(n) acts as a group of automorphisms of H. Let Γ be the subgroup of Sym(n) fixing the first k-subset of Ω. Thus Γ is isomorphic to Sym(k) × Sym($n-k$). Let P and R respectively be the characteristic matrices for the partitions determined by the orbits of Γ on k- and l-subsets of Ω. Finally let Φ be the corresponding quotient of B. It is important to note that Φ is a triangular matrix of order $(k+1) \times (k+1)$ with non-zero diagonal entries. In particular, it is invertible.

If $\gamma \in \Gamma$, let $x\gamma$ be the vector with $(x\gamma)_i = x_{i\gamma}$. We set

$$y := \frac{1}{|\Gamma|} \sum_{\gamma \in \Gamma} x\gamma .$$

As $x_1 = 1$ and $1\gamma = 1$ for all elements γ of Γ, it follows that $y \neq 0$. It is not too hard to show that $(x\gamma)^{\mathrm{T}}B = 0$ if and only if $x^{\mathrm{T}}B = 0$. This implies that $y^{\mathrm{T}}B = 0$. Now the entries of y are constant on the orbits of Γ and so there is a non-zero vector z such that $y^{\mathrm{T}} = z^{\mathrm{T}}P$. Then we have

$$0 = x^{\mathrm{T}}B = y^{\mathrm{T}}B = z^{\mathrm{T}}PB = z^{\mathrm{T}}\Phi R$$

and, since the rows of R are linearly independent, this implies that $z^{\mathrm{T}}\Phi = 0$. Since Φ is invertible this implies that $z = 0$ and so we are forced to conclude that there is no non-zero vector x satisfying $x^{\mathrm{T}}B = 0$, i.e., that the rows of B are linearly independent.

We note some simple applications of Theorem 2.4. Suppose that Ω is the edge set of a complete graph on n vertices. Then a k-subset of Ω is a graph on n vertices with k edges. The symmetric group Sym(n) acts on the $\binom{n}{2}$ elements of Ω. The orbits of k-subsets correspond to the isomorphism classes of graphs on n vertices with k edges. Since the incidence matrix for k- versus l-subsets of Ω has linearly independent rows, so does its quotient with respect to Sym(n). If $g_{n,k}$ denotes the number of isomorphism classes of graphs with n vertices and k edges, it follows that $g_{n,k} \leqslant g_{n,l}$ whenever $k \leqslant \min\{l, \binom{n}{2} - l\}$. We deduce from this that the sequence $g_{n,k}, k = 0, \ldots, \binom{n}{2}$ is unimodal. Perhaps a more significant application is the following. Let $p_{kl}(n)$ denote the number of partitions of the integer n into at most k parts, the largest of which is at most l.

Lemma 2.5. *The sequence* $p_{kl}(n), n = 0, \ldots, kl$ *is unimodal.*

Proof. We can define the *wreath product* $\Gamma = \mathrm{Sym}(l) \wr \mathrm{Sym}(k)$ to be the group acting on an $k \times l$ array R of "squares" by permuting the l squares in each row independently, and by permuting the k rows amongst themselves without changing the order of the squares in the rows. (So the order of Γ is $(l!)^k k!$.) Then $p_{kl}(n)$ is the number of orbits under Γ formed by the subsets of n squares from R, i.e., it counts the "Γ-isomorphism" classes of n-subsets of R. The lemma now follows as above. □

Lemma 2.5 is quite important and has a quite interesting history. The details of this, together with the above proof, will be found in Stanley (1982). The numbers $p_{kl}(n)$ arise in a remarkable variety of situations, occurring in particular as the coefficients in the expansion of the q-binomial coefficients $\begin{bmatrix} n \\ k \end{bmatrix}_q$ in powers of q. (For information on these see the chapter by Gessel and Stanley.) Although the sequence they form is unimodal, it is not log-concave. This means that some of the standard techniques for proving that a sequence is unimodal cannot be applied to derive Lemma 2.5.

Stanley (1985) has also used quotienting by orbits to re-derive Lovász's proof that any graph with more edges than its complement can be reconstructed from its edge-deleted subgraphs. This will be discussed, along with some generalisations, in the next section.

Recently, Wilson (1990) has determined the rank modulo p of the incidence matrix B of k-sets versus l-sets. In this paper he also gives a diagonal form for B, i.e., a diagonal matrix D such that $B = EDF$ for suitable integral matrices E and F with determinants equal to one. This is a very interesting result, but it seems fair to say that we do not yet know what its combinatorial significance is.

The theory of posets is an area where linear algebra has been effectively applied, and it would be remiss of us not to consider some examples. Let P be a poset with n elements. The incidence matrix $B = B(P)$ is the 01-matrix with ij-entry equal to 1 if and only if $i \leqslant j$ in P. We can always assume that B is an upper triangular matrix; this is equivalent to the existence of linear extensions of P. Since $i \leqslant i$, each diagonal entry of B is equal to 1, and so B is an invertible matrix. This means that there is little future in seeking to prove results about posets by rank arguments based on B. In fact we are going to work with the inverse of B.

This brings us to the *Möbius function* of P. This is the function $\mu = \mu_P$ on $P \times P$ defined by

$$\mu(i,j) := (B^{-1})_{ij}.$$

(For the basic theory of the Möbius function, expressed in a manner consistent with our approach, see Lovász 1979a, chapter 2. Another convenient reference is Aigner 1979.) Note that $\mu(i,i) = 1$ for all elements i of P. It is also not difficult to prove that $\mu(i,j) = 0$ unless $i \leqslant j$ in P. However, the key fact that we need is the following.

Lemma 2.6. *Let f be a function on the poset P. If the function f^* is defined by*

$$f^*(i) := \sum_{j \geqslant i} f(j),$$

then

$$f(i) = \sum_j \mu(i,j) f^*(j).$$

Proof. If we view f and f^* as column vectors with entries indexed by the elements of P then the first equation asserts that $f^* = Bf$. Hence $f = B^{-1}f^*$, which is equivalent to the second equation. □

The theory of the Möbius function consists of an interesting mixture of combinatorics, algebra and topology, and is very well developed. Explicit expressions for μ_P are known for many important posets. We will be making use of the following pretty result.

Lemma 2.7 (Wilf 1968, Lindström 1969). *Let P be a lattice with n elements, let f be a function on P and let F be the $n \times n$ matrix such that $(F)_{ij} = f(i \vee j)$. Then* $\det F = \prod_{i \in P} f^*(i)$, *where $f^*(i) = \sum_{j \in P} \mu(i,j) f(j)$.*

Proof. Let Φ be the diagonal matrix with ith diagonal entry equal to $f^*(i)$. Then it is easy to see that

$$(B\Phi B^{\mathrm{T}})_{ij} = \sum_{k \geqslant i,j} f^*(k) = \sum_{k \geqslant i \vee j} f^*(k)$$

and that, by th
Therefore $F = B$
$\det(F) =$
since $\det(B) = 1$. This

With the help of this le
nants. For examples, the
original paper of Wilf ment
tion complexity, see the pape
to combinatorial optimisation
result shows yet another use.

Theorem 2.8. *Let P be a lattice su*
a permutation π of the elements of P

Proof. Define a function g on P by

$$g(i) := \begin{cases} 1, & \text{if } i = 1; \\ 0, & \text{otherwise} \end{cases}$$

and let G be the matrix with ij-entry equal to ply Lemm
From Lemma 2.6 we see that $g(i) = \sum_{j \geqslant i} f(j)$ $= \mu(j,1)$. Accordi
we deduce that $\det G = \prod_{j \in P} \mu(j,1)$. Our hypo s concerning μ thus forces the conclusion that $\det G \neq 0$. The assertion of the theorem now follows from the definition of the determinant as a signed sum of products of elements of a matrix. □

Theorem 2.8 was first obtained by Dowling and Wilson (1975) using linear algebra, but not Wilf's lemma. (The above proof might even be new.) Many interesting lattices have Möbius functions that satisfy the hypothesis of the theorem. In particular, geometric lattices have this property and so have "complementing permutations" as described. From this it follows very quickly that every finite geometric lattice has at least as many hyperplanes as points. We cannot resist the following remarks in this context. Let H be a hypergraph with the property that any two distinct vertices lie in exactly one edge. Then it can be shown that the vertices and edges of H form a geometric lattice. Consequently such a hypergraph has at least as many edges as vertices. This indicates that there is a non-trivial connection between Theorem 2.8 and Fisher's inequality.

To complete this section we mention another important result in the theory of posets which has been established using linear algebra, namely the proof of Ivan Rival's conjecture on modular lattices, by Kung (1985, 1987).

3. Designs and codes

We introduce a framework which will enable us to derive some far-reaching generalisations of Fisher's inequality, and a number of other results. Our approach follows the exposition in Godsil (1993, chapters 14–16).

A *separation function* ρ on a set Ω is simply a function on $\Omega \times \Omega$ taking values in some field, the reals unless otherwise notified. If f is a real polynomial with degree r and $a \in \Omega$ then we call the mapping

$$x \rightarrow f(\rho(a,x))$$

a *zonal polynomial of degree at most* r, and denote it by f_a. We inductively define vector spaces $\mathrm{Pol}(\Omega,r)$ as follows. We set $\mathrm{Pol}(\Omega,0)$ equal to the span of the constant functions on Ω and $\mathrm{Pol}(\Omega,1)$ equal to the span of the zonal polynomials f_a, where f ranges over all real polynomials of degree at most one and a over the points in Ω. If $r > 1$ then $\mathrm{Pol}(\Omega,r)$ is the space spanned by

$$\{fg\colon\ f \in \mathrm{Pol}(\Omega,1),\ g \in \mathrm{Pol}(\Omega,r-1)\}.$$

We also define

$$\mathrm{Pol}(\Omega) = \bigcup_{r \geqslant 0} \mathrm{Pol}(\Omega,r).$$

We refer to the elements of $\mathrm{Pol}(\Omega)$ as *polynomials* on Ω, and a polynomial which lies in $\mathrm{Pol}(\Omega,r)$, but not in $\mathrm{Pol}(\Omega,r-1)$, will be said to have *degree r*. Note that if f is a polynomial of degree r on Ω and g is a polynomial of degree s then the product fg will be a polynomial of degree at most $r+s$. (Note also that $x^2+y^2+z^2$ is a polynomial of degree zero on the unit sphere in $\mathbb{R}^3$.)

A *polynomial space* consists of a set Ω, a separation function ρ on Ω and an inner product $(\cdot,\cdot)$ on $\mathrm{Pol}(\Omega)$ such that the following axioms hold:

(I) If $x,y \in \Omega$ then $\rho(x,y) = \rho(y,x)$.

(II) The dimension of $\mathrm{Pol}(\Omega,1)$ is finite.

(III) $(f,g) = (1,fg)$ for all f and g in $\mathrm{Pol}(\Omega)$.

(IV) If $f \in \mathrm{Pol}(\Omega)$ and $f(x) \geqslant 0$ for all x in Ω, then $(1,f) \geqslant 0$, with equality if and only if $f = a$.

These axioms are not very restrictive. Moreover, when Ω is finite, Axioms (II) and (IV) are redundant. We now present a number of examples. In all of the cases where Ω is finite the inner product is given by

$$(f,g) = \frac{1}{|\Omega|}\sum_{x\in\Omega} f(x)g(x).$$

(a) *The Johnson scheme* $J(n,k)$.

Here Ω is the set of all k-subsets of a set of n elements and $\rho(x,y) := |x \cap y|$. For this scheme we will usually assume implicitly that $2k \leqslant n$.

(b) *The power set* $\mathbf{2}^n$.

In this case Ω is the power set of a finite set with n elements, and $\rho(x,y) := |x \cap y|$ once again.

(c) *The Hamming scheme* $H(n,q)$.

Let Σ be an alphabet of q symbols $\{0,1,\ldots,q-1\}$. Define Ω to be the set Σ^n of all n-tuples of elements of Σ, and let $\rho(x,y)$ be the number of coordinate places in which the n-tuples x and y agree. Thus $n-\rho(x,y)$ is the Hamming distance between x and y. (We note that $H(n,2)$ and $\mathbf{2}^n$ have the same underlying set, but the functions ρ are different.) We do not require q to be prime power. The elements of $H(n,q)$ are usually called *words* over Σ.

(d) *The symmetric group* Sym(n).

We set $\Omega = \mathrm{Sym}(n)$. If x and y are elements of Ω then $\rho(x,y)$ is the number of points left fixed by the permutation $x^{-1}y$. Note that we can view Sym(n) as a subset of $H(n,n)$, and that the function ρ on Sym(n) is then just the restriction of the corresponding function in $H(n,n)$.

(e) *The Grassmann scheme* $J_q(n,k)$.

This time Ω is the set of all k-dimensional subspaces of an n-dimensional vector space over a field with q elements, and $\rho(U,V)$ is the number of 1-dimensional subspaces of $U \cap V$.

(f) *The unit sphere in* $\mathbb{R}^n$.

The set Ω is formed by the unit vectors in $\mathbb{R}^n$ and $\rho(x,y)$ is the usual inner product on $\mathbb{R}^n$. In this case the elements of Pol(Ω) are precisely the polynomials in n variables, restricted to the sphere. If f and g are two elements of Pol(Ω) then their inner product is

$$(f,g) = \int_\Omega fg \, \mathrm{d}\mu,$$

where μ is the usual measure on the sphere in $\mathbb{R}^n$, normalised so that the sphere has measure 1.

(g) *Perfect matchings in* K_{2n}.

If x and y are perfect matchings in K_{2n} then $\rho(x,y)$ is the number of edges they have in common.

Let (Ω,ρ) be a polynomial space. If Φ is a finite subset of Ω and f and g are polynomials on Ω then we define

$$(f,g)_\Phi = \frac{1}{|\Phi|} \sum_{x \in \Phi} f(x)g(x).$$

We call Φ a *t-design* in Ω if

$$(1,f)_\Phi = (1,f)$$

for all f in Pol(Ω,t). A t-design in the Johnson scheme is a simple t-design, as defined in section 2. A t-design in the Hamming scheme is the same thing as a "simple" *orthogonal array*. (These claims are not trivial; a proof of the first and an outline of a proof of the second can be found in Godsil 1988.) A t-design $\mathscr{D}$ in the power set of X can be shown to be a collection of subsets of X such that, for all $s \leqslant t$, each set of s points lies in the same number of elements of $\mathscr{D}$. For the unit sphere, our definition of a t-design is the usual one. (Delsarte et al. 1977 study t-designs on the unit sphere at some length.)

These examples show that t-designs in polynomial spaces are objects of interest, and indicate the importance of the following result.

Theorem 3.1. *Let (Ω, ρ) be a polynomial space. If Φ is a t-design in Ω then $|\Phi| \geqslant \dim(\mathrm{Pol}(\Omega, \lfloor t/2 \rfloor))$.*

Proof. Let δ_{ij} be the Kronecker delta function and let $h_1, \ldots, h_n$ be an orthonormal basis for $\mathrm{Pol}(\Omega, \lfloor t/2 \rfloor)$. (Such a basis can always be found by Gram–Schmidt orthogonalisation.) Then

$$(h_i, h_j)_\Phi = (1, h_i h_j)_\Phi = (1, h_i h_j) = (h_i, h_j) = \delta_{ij}.$$

Therefore the restrictions to Φ of the polynomials h_i form a linearly independent set of functions on Φ. Since the vector space of all functions on Φ has dimension $|\Phi|$, it follows that $n \leqslant |\Phi|$. □

For this result to be useful, we need to know the dimensions of the spaces $\mathrm{Pol}(\Omega, r)$. This is a non-trivial task, but the answer is known in many cases. (Again, see Godsil 1993 for the details.) For the Johnson scheme $J(n,k)$ we have $\dim(\mathrm{Pol}(\Omega, r)) = \binom{n}{r}$ when $r \leqslant k$.

Corollary 3.2 (Ray-Chaudhuri and Wilson 1975). *Let $\mathcal{D}$ be a $2s$-design formed from the k-subsets of an n-set, with $2k \leqslant n$. Then $\mathcal{D}$ contains at least $\binom{n}{s}$ blocks.*

If (Ω, ρ) is the Hamming scheme $H(n,q)$ then $\dim(\mathrm{Pol}(\Omega, r))$ is equal to

$$\sum_{i \leqslant r} (q-1)^i \binom{n}{i}.$$

Corollary 3.3. *Let $\mathcal{D}$ be an orthogonal array of strength $2s$ in the Hamming scheme $H(n,q)$. Then*

$$|\mathcal{D}| \geqslant \sum_{i \leqslant s} (q-1)^i \binom{n}{s}.$$

The dimension of $\dim(\mathrm{Pol}(\Omega, r))$ is the same for the power set of an n-set as it is for the Hamming scheme $H(n,2)$. For the q-Johnson scheme $\mathrm{Pol}(\Omega, r)$ has dimension $\left[{n \atop r} \right]_q$, and for the unit sphere in $\mathbb{R}^n$ it has dimension $\binom{n+r-1}{r} + \binom{n+r-2}{r-1}$. (This lower bound on the size of a spherical t-design was derived by Delsarte et al. 1977.)

A $2s$-design realising the bound of Theorem 3.1 is called a *tight* design. A tight 2-design in the Johnson scheme is better known as a *symmetric* design; such designs may be said to be rather plentiful. On the other hand it has been proved (Bannai 1977) that if $t = 2s \geqslant 10$ then there are only finitely many tight t-designs in the Johnson scheme. There is also a close connection with the theory of association schemes; we will discuss this briefly following Corollary 3.9.

Our definition of a design in a polynomial space can be extended. A *weighted t-design* on a polynomial space (Ω, ρ) is a non-negative function ϕ with finite support, S say, such that

$$\sum_{x \in S} \phi(x) f(x) = (1, f)$$

for all polynomials f in $\mathrm{Pol}(\Omega, t)$. For example, if Φ is a t-design we might take ϕ to be the function equal to $1/|\Phi|$ on the elements of Φ and zero elsewhere. A weighted design in the Johnson scheme is equivalent to a design in the usual sense of the word, with repeated blocks permitted. Theorem 3.1 can be easily extended to show that, if S is the support of a weighted t-design, then $|S| \geqslant \dim(\mathrm{Pol}(\Omega, \lfloor t/2 \rfloor))$. It can also be shown, under fairly general conditions, that a polynomial space contains weighted t-designs supported by at most $\dim(\mathrm{Pol}(\Omega, t))$ points. (See Godsil 1988, 1993.) We give a simple and direct proof of this fact for the Johnson scheme.

Lemma 3.4. *For any integers t, k and v with $t < k \leqslant v - k$, there is a k-uniform hypergraph H with at most $\binom{v}{t}$ edges that is the support of a weighted t-design.*

Proof. Let X be a fixed set of v elements, and let $B_{t,k}$ be the 01-matrix with rows indexed by the t-subsets of X, columns indexed by the k-subsets and with ij-entry equal to 1 if and only if the ith t-subset is contained in the jth k-subset. A weighted t-design corresponds to a column vector x of length $\binom{v}{k}$ with non-negative entries such that

$$B_{t,k}x = j. \tag{3.1}$$

We know that (3.1) does have non-negative solutions — $\binom{v-t}{k-t}^{-1} j$, for example. Hence, by standard results in the theory of linear programming, (3.1) has non-negative basic solutions, i.e., solution vectors supported by linearly independent sets of columns of $B = B_{t,k}$. Such a set of columns has cardinality at most $\binom{v}{t}$, since this is the number of rows of B. □

Here we should also mention Wilson's well-known proof that weighted t-(v, k, λ) designs exist whenever the obvious divisibility conditions are satisfied (Wilson 1973), which also starts with eq. (3.1).

There is another lower bound on the size of a t-design which, despite its simple proof, is very useful.

Theorem 3.5. *Let Φ be a t-design in the polynomial space (Ω, ρ). Then, for any polynomial p of degree at most t which is non-negative on Φ and any point α in Φ,*

$$|\Phi| \geqslant \frac{p(\alpha)}{(1, p)}$$

and equality holds if and only if p vanishes on $\Phi \setminus \alpha$.

Proof. Let φ be a weighted t-design and let α be a point in its support. Suppose that p is a polynomial of degree at most t on Ω, and that p is non-negative on the support of φ. Then

$$\varphi(\alpha)p(\alpha) \leqslant \sum_{x:\, \varphi(x) \neq 0} \varphi(x)p(x) = (1, p),$$

from which our bound follows immediately. □

Theorem 3.5 is a form of Delsarte's linear programming bound. (See, e.g., Delsarte et al. 1977.) The name arises because this theorem suggests the following optimization problem: choose p in $\mathrm{Pol}(\Omega, t)$ non-negative on Φ so that $p(\alpha)/(1,p)$ is maximal. This is easily expressed as a linear programming problem.

Let Δ be a set of real numbers. (In all cases of interest, it will be finite.) A *Δ-code* in a polynomial space (Ω, ρ) is a subset Φ such that

$$\{\rho(x,y)\colon\ x,y \in \Phi, x \neq y\} \subseteq \Delta.$$

We will also refer simply to codes when the set Δ is determined by the context, or is not important. We say Φ has degree d if it is a Δ-code for some set Δ of cardinality d. Many interesting problems in combinatorics are equivalent to questions concerning the maximum cardinality of Δ-codes. We have a general upper bound on the cardinality of codes, but to state this we require another definition. Suppose ρ is a separation function on a set Ω and $\Phi \subseteq \Omega$. We say Φ is *orderable* if there is linear ordering "$<$" such that, whenever $a \in \Phi$,

$$\rho(a,a) \in \{\rho(a,x)\colon\ x < a\}.$$

If Φ is an orderable subset then so is any subset of it. In all the examples of polynomial spaces we listed, Ω itself was orderable. The following result is therefore significant.

Theorem 3.6. *Let ρ be a separation function on the set Ω and let Φ be an orderable subset of Ω with degree s. Then*

$$|\Phi| \leqslant \dim(\mathrm{Pol}(\Omega, s)).$$

Proof (*We only give an outline, see* Godsil 1993, Theorem 14.4.1, *for more details*). For each a in Φ let $\Delta(a)$ be the set

$$\{\rho(a,x)\colon\ \rho(x,x) \leqslant \rho(a,a), x \neq a\}$$

and let F_a be the polynomial on Ω defined by

$$F_a(x) = \prod_{\lambda \in \Delta(a)} (\rho(a,x) - \lambda).$$

Then $F_a(b) = 0$ if $b < a$ and $F_a(a) \neq 0$. Using this it is not difficult to show that the functions F_a are linearly independent. Since they also all lie in $\mathrm{Pol}(\Omega, s)$, the result follows. □

The basic technique used in proving Theorem 3.6 is due to Koornwinder (1976). We now list some of the consequences of Theorem 3.6. A set of degree s in the unit sphere is usually called an *s-distance set.*

Corollary 3.7 (Delsarte et al. 1977). *If Φ is an s-distance subset of the unit sphere in $\mathbb{R}^n$ then* $|\Phi| \leqslant \binom{n+s-1}{s} + \binom{n+s-2}{s-1}$.

Corollary 3.8 (Ray-Chaudhuri and Wilson 1975). *Let H be a k-uniform hypergraph on v vertices and let Δ be a set of positive integers with $|\Delta| = d$. Then if H is a Δ-code, $|E(H)| \leqslant \binom{v}{d}$.*

Corollary 3.9 (Frankl and Wilson 1981). *Let H be a k-uniform hypergraph on v vertices and let Δ be set of positive integers. Suppose that Δ has d' distinct elements modulo the prime p, and none of these is congruent to k modulo p. Then if H is a Δ-code, $|E(H)| \leqslant \binom{v}{d'}$.*

Corollary 3.10 (Frankl and Wilson 1981). *Let $\mathscr{F}$ be a subset of the power set of X, where $|X| = n$. If $\mathscr{F}$ has degree s then $|\mathscr{F}| \leqslant \sum_{i\leqslant s}\binom{n}{i}$.*

More information about the above results will be found in chapters 14 and 24. The paper by Frankl and Wilson (1981) contains many significant results, one of which was recently used in Kahn and Kalai (1992) to disprove Borsuk's conjecture. (This asserted that a set of diameter one in $\mathbb{R}^d$ could always be partitioned into $d+1$ sets of diameter smaller than one. Kahn and Kalai show that at least $(1.1)^{\sqrt{d}}$ such sets may be required.) Many of the polynomial spaces we have mentioned are association schemes. Delsarte (1973) showed how to define designs and codes in association schemes; where these concepts overlap ours, they agree. Further information will be found in chapter 15.

A number of interesting results of coding type have been proved using exterior algebra. The basic example is the following, which is a slight extension of a result due to Bollobás (1965). The version stated, and its proof, are due to Lovász (1977).

Theorem 3.11. *Suppose that $A_1,\ldots,A_m$ are r-element subsets of a set X, and $B_1,\ldots,B_m$ are s-element subsets of X. If $A_i \cap B_i = \emptyset$ for all i and $A_i \cap B_j \neq \emptyset$ whenever $i < j$, then $m \leqslant \binom{r+s}{s}$.*

Proof. Let f be a mapping from X into $V = \mathbb{R}^{r+s}$ such the image of any set of $r+s$ distinct points from X is linearly independent. (We could assume that f maps each element of X to a vector of the form

$$(1,t,\ldots,t^{r+s-1}).$$

It is a simple exercise to show that this works, provided only that we use distinct values of the parameter t for distinct elements of X.)

To any set S of elements of X we associate the wedge product

$$\bigwedge_{x\in S} f(x)$$

and we denote this by $\omega(S)$. (This product does dep
the multiplication is performed, but a change of orde
sign, and this will cause no problems.) Observe that t
dimension $\binom{r+s}{|S|}$, and it is non-zero if and only if the
linearly independent. If T is a second subset of X the
and only if $f(S \cup T)$ spans a subspace of V with dim

The m vectors $\omega(A_i)$ lie in a vector space of dimension $\binom{r+s}{s}$; if we can show they are linearly independent then the theorem is proved. Suppose we have scalars c_i such that

$$\sum_{i=1}^{m} c_i\, \omega(A_i) = 0. \tag{3.2}$$

Let j be the greatest index such that $c_j \neq 0$. Since $B_j \cap A_i$ is nonempty for all i less than j, we have $\omega(A_i) \wedge \omega(B_j) = 0$ if $i < j$. Since $B_j \cap A_j = \emptyset$, it follows that $f(A_j \cup B_j)$ is a linearly independent set. Hence $\omega(A_j) \wedge \omega(B_j) \neq 0$. Therefore (3.2) yields

$$\begin{aligned} 0 &= \sum_{i\ 1}^{m} c_i\, \omega(A_i) \wedge \omega(B_j) \\ &= \sum_{i \geqslant j} c_i\, \omega(A_i) \wedge \omega(B_j) \\ &= c_j\, \omega(A_j) \wedge \omega(B_j). \end{aligned}$$

But this implies that $c_j = 0$, and this forces us to conclude that the vectors $\omega(A_i)$ are linearly independent. Hence $m \leqslant \binom{r+s}{s}$. □

A subspace U of $V = \mathbb{R}^{r+s}$ with basis $v_1, \ldots, v_m$ can be represented by the vector $\bigwedge_i v_i$. Hence the argument used above yields the following result.

Lemma 3.12 (Lovász 1977). *If we are given r-dimensional subspaces* $U_i, \ldots, U_m$ *and s-dimensional subspaces* $W_1, \ldots, W_m$ *of* $V = \mathbb{R}^{r+s}$ *such that* $U_i \cap W_j \neq 0$ *if* $i < j$ *and* $U_i \cap W_i = 0$ *then* $m \leqslant \binom{r+s}{s}$.

The theorem itself is a consequence of this lemma, together with the observation that there is an injection of X into V which maps all subsets with cardinality at most $r+s$ onto independent sets. In fact the lemma holds independently of the dimension of V. For suppose we have subspaces U_i and W_j as described in a vector space V, where $\dim(V) > r+s$. Since we can extend the field we are working over if necessary, there is no loss in assuming it is infinite. Choose a subspace V_0 of V with codimension $r+s$ in general position with respect to the subspaces U_i and W_j, and let ϕ denote the mapping onto the quotient space V/V_0. Then $\dim(U_i \cap W_j) = \dim(\phi(U_i) \cap \phi(W_j))$ for all i and j and we can now apply the lemma to the subspaces $\phi(U_i)$ and $\phi(W_j)$, $1 \leqslant i, j \leqslant m$, of the vector space V/V_0. (One consequence of this is that Theorem 3.11 actually holds if the A_i and B_j are flats of rank r and s respectively in a linear matroid.)

More examples of the use of exterior algebra will be found in Lovász (1977, 1979c) and Alon (1985). One possible source for background on exterior algebra is Northcott (1984), but any book on multilinear algebra would suffice for what we have used.

4. Null designs

Let V be a fixed set with v elements. A function f on the subsets of V is a *null design of strength* t (or a null t-design) if, for each subset τ of V with at most t elements,

$$\sum_{\beta \supseteq \tau} f(\beta) = 0. \tag{4.1}$$

If U is a subset of V then the restriction of f to the subsets of U is not, in general, a null design of strength t on U. However, there is an easy way to construct such a function from f, due to Frankl and Pach (1983), that we now describe.

Given any function f on the subsets of V, define the function f^* by setting

$$f^*(\tau) := \sum_{\beta \supseteq \tau} f(\beta). \tag{4.2}$$

Then f is a null t-design if and only if f^* vanishes on the subsets of V with at most t elements. Also f can be recovered from f^* by Möbius inversion thus:

$$f(\beta) = \sum_{\tau \supseteq \beta} (-1)^{|\tau - \beta|} f^*(\tau). \tag{4.3}$$

Consequently we can construct a null t-design on the subset U of V as follows.

(a) Choose a null t-design f on V.

(b) Compute the transform f^* as in (4.2) above.

(c) Apply Möbius inversion on the subsets of U (as in (4.3)) to the restriction $(f^*) \upharpoonright U$ of f^* to U.

Let us denote the resulting function by f_U. We can view it as a null design on V by the simple expedient of defining it to be zero on any subset of V not contained in U.

There is a possibility that f_U may be identically zero, but this will not happen unless f^* vanishes on all subsets of U. We have

$$\begin{aligned} f_U(\alpha) &= \sum_{\alpha \subseteq \beta \subseteq U} (-1)^{|\beta - \alpha|} f^*(\beta) \\ &= \sum_{\alpha \subseteq \beta \subseteq U} (-1)^{|\beta - \alpha|} \sum_{\gamma \supseteq \beta} f(\gamma) \\ &= \sum_{\gamma \subseteq V} f(\gamma) \sum_{\alpha \subseteq \beta \subseteq \gamma \cap U} (-1)^{|\beta - \alpha|} \\ &= \sum_{\gamma \cap U = \alpha} f(\gamma) \end{aligned} \tag{4.4}$$

which provides a useful alternative definition of f_U. One consequence of (4.4) is that if $f_U(\alpha) \neq 0$ then $f(\gamma) \neq 0$ for some subset γ of V such that $U \cap \gamma = \alpha$. We also obtain the following result.

Lemma 4.1. *Let f be a null design of strength t on the set V and let U be a minimal subset of V such that $f^*(U) \neq 0$. Then if $\alpha \subseteq U$,*

$$f_U(\alpha) = (-1)^{|U \setminus \alpha|} f^*(U).$$

Proof. This follows immediately from the definition of f_U. □

Corollary 4.2. *Any non-zero null design of strength t on the set V assumes a non-zero value on at least 2^{t+1} subsets of V.*

Proof. Let U be a minimal subset of V such that $f^*(U) \neq 0$. Since f has strength t, the cardinality of U is at least $t+1$. By the lemma, f_U is non-zero on each subset of U and so, by the remark above, for each subset α of U, there must be a subset γ of V such that $\gamma \cap U = \alpha$ and $f(\gamma) \neq 0$. This supplies us with $2^{|U|}$ distinct elements of V on which f is non-zero. □

Let G be the incidence matrix for the subsets of a v-set with cardinality at most t, versus all subsets of the same v-set. Then a null t-design can be viewed as an element of the null-space of G, and so Corollary 4.2 can viewed as determining the minimum distance of a code over the rationals. If we had worked modulo 2 we would have obtained a *Reed–Muller* code. The minimum distance of these codes has been determined, and is given in most textbooks on coding theory. (See chapter 16 in this Handbook or, for example, MacWilliams and Sloane 1978, chapter 13.) The arguments used to determine this minimum distance actually suffice to determine the minimum distance over the rationals. Hence we may view the above corollary as a translation of a known result. Corollary 4.2 is also derived, in another context, in Anstee (1985, Proposition 2.5). We now present some applications of this machinery.

Lemma 4.3 (Frankl and Pach 1983). *If H_1 and H_2 are two distinct t-designs with the same vertex set then the symmetric difference of their edge sets contains at least 2^{t+1} edges.*

Proof. Let χ_1 and χ_2 be the respective characteristic vectors of H_1 and H_2. Then it is not difficult to check that $\chi_1 - \chi_2$ is a null design of strength t. By Corollary 4.2 it must have at least 2^{t+1} non-zero entries. □

Our next application of Corollary 4.2 requires some further preliminaries. A hypergraph H_1 is an *edge-reconstruction* of the hypergraph H_2 if there is a bijection ϕ from $E(H_1)$ to $E(H_2)$ such that, for each edge e in H_1, the edge-deleted hypergraph $H_1 \setminus e$ is isomorphic to $H_2 \setminus \phi(e)$. We say that a hypergraph H is *edge-reconstructible* if any hypergraph that is an edge reconstruction of H is isomorphic to it. Thus we can say that a hypergraph is edge reconstructible if it is determined by the collection of its edge deleted hypergraphs. The edge reconstruction conjecture for graphs asserts that all graphs with at least four edges are edge-reconstructible. Bondy and Hemminger (1977) provide an excellent, if slightly dated, survey of progress on the reconstruction problem.

A hypergraph is *s-edge reconstructible* if it is determined by the collection of $\binom{e}{s}$ hypergraphs obtained by deleting, in turn, each set of s edges from it. The next result generalises the result of Müller (1977) on edge reconstruction of graphs.

Lemma 4.4. *Let H be a hypergraph with v vertices and e edges. If $2^{e-s} > v!$ then H is s-edge reconstructible.*

Proof. Assume by way of contradiction that H_1 and H_2 are two non-isomorphic hypergraphs with e edges, and the same collection of s-edge deleted hypergraphs. There is no loss of generality in assuming that H_1 and H_2 have the same vertex set V. We view a hypergraph with vertex set V as a subset of the power set $\mathbf{2}^V$ of V. If $i = 1$ or 2, let χ_i be the function on the $\mathbf{2}^V$ defined by

$$\chi_i(F) = \begin{cases} 1, & \text{if } F \cong H_i; \\ 0, & \text{otherwise.} \end{cases}$$

I claim that the function

$$\chi := |\mathrm{Aut}(H_1)|\chi_1 - |\mathrm{Aut}(H_2)|\chi_2$$

is a null design with strength $e - s$ on $\mathbf{2}^V$. For if L is any hypergraph with vertex set V and $i = 1$ or 2 then

$$\sum_{F \supseteq L} |\mathrm{Aut}(H_i)|\chi_i(F)$$

is equal to the number of permutations τ of V such that the image of H_i under τ contains L, and this is in turn equal to the number of sub-hypergraphs of H_i isomorhic to L. The claim that χ is a null design with strength $e - s$ is consequently a restatement of the hypothesis that H_1 and H_2 have the same s-edge deleted sub-hypergraphs.

It follows that χ must take non-zero values on at least 2^{e-s+1} hypergraphs. But $|\mathrm{Aut}(H_i)|\chi_i$ is equal to 1 on each of $|\mathrm{Sym}(V)|/|\mathrm{Aut}(H_i)|$ hypergraphs with vertex set V that are isomorphic to H_i ($i = 1,2$), and is equal to zero on all others. Thus it takes non-zero values on at most $2|\mathrm{Sym}(V)| = 2v!$ hypergraphs. This means that we must have $2^{e-s} \leqslant v!$ □

Let B be the incidence matrix of hypergraphs with $e - s$ edges versus hypergraphs with e edges (and all having vertex set V). If χ is a non-zero null design with strength $e - s$ then $B\chi = 0$. Hence the columns of B must be linearly dependent. From Theorem 2.4 it follows that in this case B must have more rows than columns. So if χ exists as described then

$$\binom{2^v}{e-s} > \binom{2^v}{e},$$

which implies that $e - s < 2^v - e$. Thus we have deduced the following.

Lemma 4.5. *Let H be a hypergraph with v vertices and e edges. If $2e \geqslant 2^v + s$ then H is s-edge reconstructible.*

When $s = 1$ this result was first proved in Lovász (1972), using an inclusion-exclusion argument. A proof using a form of quotient argument was subsequently presented in Stanley (1985). The argument just used is easily modified to prove that a k-uniform hypergraph on v vertices with e edges is s-edge reconstructible if $2e \geqslant \binom{v}{k} + s$. On the other hand Lemma 4.4 holds as stated for k-uniform hypergraphs. For graphs, the analogues of Lemmas 4.4 and 4.5 were first proved in Godsil et al. (1987).

So far, all our applications of the theory of null designs have used only Corollary 4.2. We now give an example where Lemma 4.1 is used. A hypergraph is *k-chromatic* if we can partition its vertex set into k classes such that no edge is a subset of any one of the classes. It is *critically* k-chromatic if it is k-chromatic and each of the subgraphs obtained by deleting one edge from it is $(k-1)$-chromatic. Thus the cycle on five vertices is an example of a critically 3-chromatic 2-uniform hypergraph. The result we are about to prove, due to Lovász (1976), asserts that any critically 3-chromatic k-uniform hypergraph with v vertices has at most $\binom{v}{k}$ edges. This is an immediate byproduct of the following.

Lemma 4.6 (Lovász 1976). *Let H be a critically 3-chromatic k-uniform hypergraph with vertex set V and let $B = B_{k-1}(H)$ be the incidence matrix for the $(k-1)$ subsets of V versus the edges of H. Then the columns of B are linearly independent.*

Proof. Assume by way of contradiction that the columns of B are linearly dependent. Then there is a null design f of strength $(k-1)$ on V that is supported by the edges of H. Thus f^*, as defined by eq. (4.2) above, vanishes on all subsets of V with fewer than k elements. Since f itself vanishes on all subsets of V with more than k elements, it follows from (4.2) that $f = f^*$.

Now let (X, Y) be any partition of V into two classes. Then, from (4.4) we have

$$f_X(\emptyset) = \sum_{\gamma \cap X = \emptyset} f(\gamma).$$

Since $f = f^*$ it follows from this that the above sum is equal to $\sum_{\gamma \subseteq Y} f^*(\gamma)$ and, given that $f^*(\gamma) \neq 0$ only when $\gamma \in E(H)$, we thus deduce that

$$f_X(\emptyset) = (-1)^k f_Y(\emptyset).$$

Using (4.4) once more we obtain

$$\sum_{\beta \cap X = \emptyset} f(\beta) = (-1)^k \sum_{\beta \cap Y = \emptyset} f(\beta). \tag{4.5}$$

To complete the proof we choose an edge α of H such that $f(\alpha) \neq 0$ and take (X, Y) to be a 2-colouring of $H \setminus \alpha$. Then α is the unique edge of H contained in one of the sets X and Y. This implies that one side of (4.5) is zero, but the other is not. Accordingly f cannot exist as described, and so the columns of $B_{k-1}(H)$ are linearly independent. □

The above proof is no simpler than the original, and differs from it only in the argument used to derive (4.5). However, it does show how the available information on null designs can be used. There is a closely related result due to Seymour.

Lemma 4.7 (Seymour 1974). *The rows of the incidence matrix of a critically 3-chromatic hypergraph are linearly independent over* $\mathbb{R}$.

Proof. Let H be a critically 3-chromatic hypergraph with incidence matrix B. Assume by way of contradiction that there is a non-zero vector y such that $y^{\mathrm{T}}B = 0$. The hypergraph induced by the vertices i such that $y_i = 0$ is 2-colourable. Assume that it has been coloured blue and red. Extend this to H by colouring the vertices j such that $y_j > 0$ with blue, and the remaining vertices red. If b is a column of B then $y^{\mathrm{T}}b = 0$. Hence either $y_i = 0$ for all vertices i in the edge corresponding to b, or else y is positive on one vertex of this edge and negative on another. This shows that our colouring of the vertices of H is a proper 2-colouring, which contradicts our choice of H. □

This proof is interesting in that it depends on the fact that $\mathbb{R}$ is an ordered field. No other example of this comes to mind. The Fano plane shows that the result is not valid over finite fields.

We remark finally that there is a close connection between the theory of null designs and the representation theory of the symmetric group. The key to this is that we may identify a k-subset of a v-set with a "tabloid" having two rows, of size $v - k$ and k. (As ever, we assume $2k \leqslant v$.) Then the null designs with minimum support constructed in Frankl and Pach (1983) can be viewed as "polytabloids", which span a Specht module for the symmetric group. For more information on the latter see, e.g., James (1978, chapter 4).

5. Walks in graphs

In the previous sections our emphasis has been on design theory, but from now it will be on graphs (and directed graphs). We begin by establishing some notation. An edge $\{u, v\}$ in a graph will be regarded as being formed from the two *arcs* (u, v) and (v, u). (This usage of the term "arc" is also standard in other situations, e.g., when discussing automorphism groups of graphs.) Hence we may, when convenient, view a graph as simply a special type of directed graph. If D is a directed graph with vertex set V then its *adjacency matrix* $A(D)$ is the matrix with rows and columns indexed by V, and with uv-entry equal to the number of arcs in D from u to v. (Our directed graphs may have loops and/or parallel arcs, however our graphs will always be simple.) Note that isomorphic directed graphs will not in general have the same adjacency matrices but, as will become apparent, this is never the source of any problems.

A *walk* in a directed graph is a sequence

$$v_0, e_1, v_1, \ldots, v_{n-1}, e_n, v_n$$

formed alternately of vertices and arcs, such that e_i is the arc (v_{i-1}, v_i). The *length* of the above walk is n. We explicitly permit walks of length zero; there is one such walk for each vertex. A walk that starts and finishes at the same vertex is called *closed*. All walks, even in undirected graphs, are directed objects. The basic result concerning walks can now be stated.

Lemma 5.1. *Let D be a directed graph with adjacency matrix A. If u and v are vertices of D then $(A^k)_{uv}$ is equal to the number of walks of length k in D that start at u and finish at v.*

The proof of this result is a routine induction argument, based on the observation that $A^k = AA^{k-1}$. One consequence of this result is that $\operatorname{tr} A^k$ is equal to the number of closed walks in D with length k. (And since $A^0 = I$, we thus reconfirm that there is one closed walk of length zero on each vertex of D.) We note also that if D is a graph then $\operatorname{tr} A = 0$, $\operatorname{tr} A^2$ equals twice the number of edges in D and $\operatorname{tr} A^3$ is equal to six times the number of 3-cycles. Given the existence of fast algorithms for matrix multiplication, the last observation leads to the most efficient known algorithm for detecting a triangle. This also works when D is directed, provided we first delete all the loops from it. (This approach to finding 3-cycles has occurred independently to a number of people, so I remain silent on the question of its attribution. The efficiency of such a "non-combinatorial" algorithm is undoubtedly a source of annoyance in many quarters.)

The most effective way to study walks in graphs is by using generating functions. To describe this we first need another round of definitions. Let D be a directed graph with adjacency matrix A. The *walk generating function* of D is

$$W(D,x) := (I - xA)^{-1} = \sum_{k\geqslant 0} x^k A^k.$$

Thus $W(D,x)$ is a formal power series with coefficients in a ring of matrices. The uv-entry of $W(D,x)$ will be written as $W_{uv}(D,x)$ and the trace of $W(D,x)$ will be denoted by $C(D,x)$. (As we have no intention of ever setting x equal to a real or complex number in one of these series, the reader should put all thoughts of convergence from her or his mind.) The *characteristic polynomial* $\det(xI - A)$ of A will be denoted by $\phi(D,x)$ and referred to as the characteristic polynomial of D. If $u \in V(D)$ then $D\setminus u$ is the directed graph obtained by removing u, together with all its attendant arcs. Convenient references for background information on adjacency matrices and related topics are Biggs (1993) and Cvetković et al. (1980). Walk generating functions are studied at some length in Godsil (1993, chapter 4).

Lemma 5.2. *Let u be a vertex in the directed graph D. Then*

$$x^{-1}W_{uu}(D,x^{-1}) = \phi(D\setminus u,x)/\phi(D,x).$$

Proof. Let B be the adjacency matrix of $D\setminus u$. From Cramer's rule and the definition of $W(D,x)$, we see that $W_{uu}(D,x) = \det(I - xB)/\det(I - xA)$. (Remark: the two identity matrices I in this quotient have different orders. We will frequently be found guilty of this abuse of notation.) If $n = |V(D)|$ then

$$\det(I - xA) = x^n \det(x^{-1}I - A) = x^n\phi(D,x^{-1})$$

and similarly $\det(xI - B) = x^{n-1}\phi(D\setminus u, x^{-1})$. The lemma follows immediately. □

The above lemma provides an explicit expression for the diagonal entries of $W(D,x)$. We derive some analogous formulas for the off-diagonal elements later. We note one simple but useful property of the characteristic polynomial. For the proof see, for example Cvetković et al. (1980, Theorem 2.14) or Godsil (1993, Theorem 2.1.5(c)).

Lemma 5.3. *For any directed graph D,*

$$\phi'(D,x) = \sum_{u\in V(D)} \phi(D\setminus u, x).$$

As an immediate consequence of Lemmas 5.2 and 5.3, we infer that

$$x^{-1}C(D,x^{-1}) = \phi'(D,x)/\phi(D,x). \tag{5.1}$$

This shows that the characteristic polynomial and the closed walk generating function of a directed graph provide the same information. If we multiply both sides of (1) by $\phi(D,x)$ and then equate coefficients, we recover a system of equations connecting the sums of the powers of the zeros of $\phi(D,x)$ with its coefficients.

The concept of quotients, as introduced in section 2, can be applied very usefully to graphs and directed graphs. It was first studied by H. Sachs; a discussion of it from his point of view is presented in Cvetković et al. (1980, chapter 4). Here we will only consider quotients of graphs, a more extensive treatment of this topic is given in Godsil (1993, chapter 5). One definition is necessary. If G is a graph then a partition π of $V(G)$ will be called *equitable* if the pair of partitions (π,π) is equitable in the sense used in section 2. We have the following.

Lemma 5.4. *Let G be a graph and let π be a partition of $V(G)$ with characteristic matrix P. Then π is equitable if and only if there is a matrix Φ such that $PA(G) = \Phi P$.*

Here Φ is a square matrix with rows and columns indexed by the cells of π and with $(\Phi)_{ij}$ equal to the number of arcs that start at a vertex in cell i and finish on a given vertex in cell j. Thus if G is Petersen's graph, u is a fixed vertex in G and π is the partition of $V(G)$ induced by the distance in G from u then

$$\Phi = \begin{pmatrix} 0 & 1 & 0 \\ 3 & 0 & 1 \\ 0 & 2 & 2 \end{pmatrix},$$

which illustrates that Φ need not be symmetric. We shall find it convenient to view Φ as the adjacency matrix of a directed graph with the cells of π as its vertices. This directed graph will be denoted by G/π. The following result can now be derived in a routine fashion.

Lemma 5.5. *Let π be an equitable partition of the graph G and set $\Phi = A(G/\pi)$. Then $(\Phi^k)_{ij}$ is equal to the number of walks of length k in G/π that start in cell i and finish at a specified vertex in cell j.*

The discrete partition, with each cell a singleton, is always equitable. Consequently Lemma 5.5 is a generalisation of the better known Lemma 5.1. The last two results provide all the information on quotients that we need.

One consequence of Lemma 5.4 is that the characteristic polynomial of G/π divides that of G. To see this note first that if U is an invariant subspace for A then we have $PU = PAU = \Phi PU$, showing that PU is an invariant subspace for Φ. From this, and the fact that the rows of P are linearly independent, it can be shown that the characteristic polynomial of Φ divides that of A. In one important case we can compute $\phi(G,x)$ from $\phi(G/\pi,x)$.

Theorem 5.6. *Let G be a graph with n vertices, let u be a vertex in G and let π be an equitable partition of G in which $\{u\}$ is a cell. Then if $\phi(G\setminus v,x)$ is the same for all vertices v in G and $H = G/\pi$,*

$$\phi'(G,x)/\phi(G,x) = n\phi(H\setminus\{u\},x)/\phi(H,x). \tag{5.2}$$

Proof. Let $C_1,\ldots,C_r$ be the cells of π and denote the corresponding vertices of H by $1,\ldots,r$. Assume that $C_1 = \{u\}$. From Lemma 5.5 we see that if the vertex v of G is in cell C_i then $W_{uv}(G,x) = W_{1i}(H,x)$. The result now follows from Lemmas 5.2 and 5.3. □

It is not difficult to show that, when $\{u\}$ is a cell π, $(\Phi^k)_{i1}/(\Phi^k)_{1i} = |C_i|$. Thus, under the hypotheses of Theorem 5.6, we can compute $\phi(G,x)$ given $\Phi = A(H)$. The most obvious case where this result can be applied is when Aut(G) is vertex transitive and π is the partition of $V(G)$ formed by the orbits of a subgroup of Aut(G) that fixes the vertex u. The next result is one of the most important applications of the theory we have described.

Corollary 5.7. *Under the hypotheses of Theorem 5.6, the numerators in the partial fraction expansion of $n\phi(H\setminus\{u\},x)/\phi(H,x)$ are the multiplicities of the zeros of $\phi(G,x)$.*

Proof. This is a well-known property of the partial fraction expansion of $\mu'(x)/\mu(x)$, for any polynomial $\mu(x)$. □

Corollary 5.7 thus provides a feasibility condition that a digraph Δ must satisfy to occur as the quotient with respect to an equitable partition π of a graph G, for which the conditions of Theorem 5.6 hold. This condition can be formulated in a number of different ways, and is often referred to as the "eigenvalue method". The key idea is that the multiplicities of the eigenvalues of $A(G)$ can be determined from a fairly limited amount of information. There are surprisingly many situations where this is useful. The "classical" application is in demonstrating the non-existence of classes of, or individual, distance-regular graphs. The most well-known, and earliest example, is provided by the work of Hoffman and Singleton (1960) on Moore graphs of diameter two and three. (A convenient description of their work, and more recent generalisations, will be found in Biggs 1993.) For another application we mention the proof of the fact that finite projective planes cannot have a null polarity, as presented in Hughes and Piper (1973), and the generalisation of this result to the so-called "friendship theorem". (For more details and further references, see Cameron and Van Lint 1991, p. 45.) This method has also recently been applied in model theory (Evans 1986), albeit at a point where the distinction between this subject and finite geometry is hard to discern. Finally, McKay (1979) has used Theorem 5.6 and Corollary 5.7 to determine, with the aid of a computer, all vertex-transitive graphs with fewer than 20 vertices.

Our approach to Corollary 5.7 is not the standard one, which is based on computations with the eigenvectors of $\Phi = A(\Delta)$, and places much more restrictive conditions on G (namely that it should be a distance-regular graph). An accessible discussion from this viewpoint is presented in Biggs (1993). A detailed exposition along the lines taken above will be found in Godsil and McKay (1980).

We are now going to derive some information about the off-diagonal elements of $W(D,x)$. The *adjugate* of $xI - A$, i.e., the transpose of its matrix of cofactors, will be denoted by $\Psi(A,x)$. The most important property of Ψ is that

$$\Psi(A,x)(xI - A) = \det(xI - A)I.$$

If A is the adjacency matrix of the directed graph D then $(\Psi(A,x))_{ii}$ is equal to $\phi(D \setminus i, x)$. In this case we denote the ij-entry of $\Psi(A,x)$ by $\phi_{ij}(D,x)$. It is easy to show that

$$x^{-1}W_{uv}(D,x) = \phi_{uv}(D,x)/\phi(D,x).$$

If A is an $n \times n$ matrix and $U \subseteq \{1,\ldots,n\}$, we denote by $\Psi_U(A,x)$ the (square) submatrix of Ψ with rows and columns indexed by the elements of U. We use $A \setminus U$ to denote the matrix obtained by deleting the rows and columns indexed by U. We need the following crucial result, the combinatorial significance of which first seems to have been noted by Tutte (1947, 1979).

Lemma 5.8 (Jacobi 1833). *If A is an $n \times n$ matrix and U is a subset of $\{1,\ldots,n\}$ with m elements then*

$$\det \Psi_U(A,x) = (\det(xI - A))^{m-1} \det(xI - (A \setminus U)).$$

Proof. We may assume without loss that $U = \{1,\ldots,m\}$. Let M be the matrix obtained by replacing the first m columns of the $n \times n$ identity matrix with the corresponding columns of $\Psi(A,x)$. Then the product $(xI - A)M$ has the form

$$\begin{pmatrix} \det(xI - A)\, I_m & 0 \\ * & xI_{n-m} - (A \setminus U) \end{pmatrix}, \tag{5.3}$$

where the diagonal blocks are square (and the details of the sub-diagonal block are irrelevant). Now $\det M = \det \Psi_U(A,x)$ and so we have

$$\begin{aligned}\phi(A,x)\det \Psi_U(A,x) &= \det\big((xI-A)M\big)\\ &= (\det(xI-A))^m \det(xI-(A\setminus U)).\end{aligned}$$

(The last term is just the determinant of the matrix in (5.3).) This equation immediately yields the lemma. □

Lemma 5.8 is in fact a classical result, best described as well forgotten. It is sometimes referred to as "Jacobi's identity", which is not a particularly useful identifier. We will only be using it when $|U| = 2$. For ease of reference we restate this case in a modified form.

Corollary 5.9. *Let G be a directed graph with vertices i and j. Then*

$$\phi_{ij}(G,x)\phi_{ji}(G,x) = \phi(G\setminus u,x)\phi(G\setminus v,x) - \phi(G,x)\phi(G\setminus\{i,j\},x).$$

When G is a graph, $\phi_{ij}(G,x) = \phi_{ji}(G,x)$ and so Corollary 5.9 implies that

$$\phi_{ij}(G,x) = \sqrt{\phi(G\setminus u,x)\,\phi(G\setminus v,x) - \phi(G,x)\,\phi(G\setminus\{i,j\},x)} \tag{5.4}$$

It might appear that the sign of $\phi_{ij}(G,x)$ is not determined by this expression, but we know that the rational function $\phi_{ij}(G,x)/\phi(G,x)$ has non-negative coefficients when expanded as a series in x^{-1}. This implies that the leading term of $\phi_{ij}(G,x)$ is always positive.

A very nice application of eq. (5.4) to graph reconstruction was found by Tutte.

Theorem 5.10 (Tutte 1979). *If the characteristic polynomial of the graph G is irreducible over the rationals then G is vertex-reconstructible.*

Proof. Let the vertex set of G be $\{1,\dots,n\}$ and suppose $\phi(G,x)$ is irreducible. We prove that for any two distinct vertices i and j of G, the polynomial $\phi(G\setminus ij,x)$ is determined by $\phi(G,x)$, $\phi(G\setminus i,x)$ and $\phi(G\setminus j,x)$. We have

$$\phi(G\setminus i,x)\,\phi(G\setminus j,x) - \phi(G,x)\,\phi(G\setminus ij,x) = \phi_{ij}(G,x)^2. \tag{5.5}$$

Now suppose that η is a polynomial such that

$$\phi(G\setminus i,x)\,\phi(G\setminus j,x) - \phi(G,x)\eta = \sigma^2 \tag{5.6}$$

for some polynomial σ of degree at most $n-2$. Then, subtracting (5.5) from (5.6), we obtain

$$\phi(G,x)\,(\phi(G\setminus ij,x) - \eta) = \phi_{ij}(G,x)^2 - \sigma^2.$$

The right side of this equation is the product of two polynomials, each of degree at most $n-2$. Since this product is divisible by $\phi(G,x)$, which is irreducible of degree n, we are forced to conclude that $\eta = \phi(G\setminus ij,x)$. This proves our claim.

As noted in the proof of Lemma 5.3, if H has m vertices then the coefficient of x^{m-2} in $\phi(H,x)$ is equal to -1 times the number of edges in H. So, given $\phi(G)$, $\phi(G\backslash i)$, $\phi(G\backslash j)$ and $\phi(G\backslash\{i,j\})$ we can determine the number of edges joining i to j, i.e., whether or not i and j are adjacent. Therefore when $\phi(G)$ is irreducible, the first three of these polynomials determine whether i and j are adjacent.

To complete the proof we now recall that in Tutte (1979) it is shown that the characteristic polynomial of a graph G is determined by the collection of vertex-deleted subgraphs of G. Hence G is vertex-reconstructible when $\phi(G)$ is irreducible. □

The above proof still works if $\phi(G)$ is not irreducible, but instead has an irreducible factor of degree $n-1$. For another variation, suppose that $\phi(G\backslash 1)$ is irreducible. An argument similar to the one above shows then that $\phi(G)$, $\phi(G\backslash 1)$ and $\phi(G\backslash\{1,i\})$ determine $\phi(G\backslash i)$. From this it follows again that G is vertex-reconstructible. This result was first proved, in apparently greater generality, in Yuan (1982). See also Godsil and McKay (1981).

There are close connections between the theory of matchings in graphs and the topics we are discussing. To describe this we require some more notation. A *k-matching* in a graph is a set of k disjoint edges, no two of which have a vertex in common. The number of k-matchings in the graph G will be denoted by $p(G,k)$. We call

$$\mu(G,x) := \sum_k (-1)^k p(G,k)x^{n-2k}$$

the *matchings polynomial* of G. The task of computing this polynomial for a given graph is NP-hard (or, more precisely, #P-complete), since the constant term of $\mu(G,x)$ counts the number of perfect matchings in G and counting the number of perfect matchings in bipartite graphs is equivalent in complexity to determining the permanent of 01-matrices. From Valiant (1979), we know that the latter is NP-hard. One consequence of this is that, unless P=NP, there is no easy way of computing $\mu(G,x)$.

Thus the matchings polynomial is in this regard a more intractable object than the characteristic polynomial of a graph. Nonetheless, it is known that G is a forest if and only if $\mu(G,x) = \phi(G,x)$ and there are also some simple recurrences that enable us to compute the matchings polynomials of small graphs with some facility. The matchings polynomials of bipartite graphs are essentially the same as "rook polynomials". (For information on rook polynomials see Riordan 1958. For the matchings polynomial see Heilmann and Lieb 1972, Farrell 1979, Godsil and Gutman 1981, and Godsil 1981b, 1993, chapters 1 and 6.)

An unexpected property of the matchings polynomial is that all its zeros are real. The first, second and third proofs of this are to be found in the above-mentioned paper of Heilmann and Lieb. For a combinatorialist this is perhaps not the easiest paper to read, and it is probably a non-trivial task even to locate all three of the proofs just referred to.) A fourth proof will follow from the next result. The fact that the zeros are real is not without combinatorial significance. It implies, for example, that the sequence formed by the numbers $p(G,k)$ $(k=0,1,\ldots)$ is log-concave. (This was noted by Heilmann and Lieb.) Another consequence is that, in many cases of interest, the number of edges in a randomly chosen matching has exactly k edges is asymptotically normally distributed. (See Godsil 1981a.)

Theorem 5.11 (Godsil 1981b). *Let G be a graph and let u be a vertex in G. Let $T = T(G,u)$ be the tree with the paths in G that start at u as its vertices, and with two such paths adjacent if and only if one is a maximal subpath of the other. Then*

$$\frac{\mu(G\setminus u,x)}{\mu(G,x)} = \frac{\mu(T\setminus u,x)}{\mu(T,x)}.$$

(In the right side of the above identity, u denotes the one-vertex path consisting of u itself.) As we remarked above, when H is a forest we have $\mu(H,x) = \phi(H,x)$. So from Theorem 5.11 we deduce that all zeros and poles of the rational function $\mu(G\setminus u,x)/\mu(G,x)$ are real. A trivial induction argument on the number of vertices in G now yields the conclusion that all the zeros of $\mu(G,x)$ are real. Another consequence of Theorem 5.11 is that $\mu(G\setminus u,x)/\mu(G,x)$ is essentially a generating function for a class of walks in G. (This because the right-hand side can be written as $\phi(T\setminus u,x)/\phi(T,x)$ and this is "essentially" a generating function, by Lemma 5.2.)

Another connection between linear algebra and the theory of matchings is provided by Pfaffians. We discuss this briefly. Let $A = (a_{ij})$ be a skew-symmetric $n \times n$ matrix, i.e., $A^{\mathrm{T}} = -A$. let $\mathcal{F}(n)$ be the set of permutations π of $\{1,\ldots,n\}$ such that all cycles of π have even length. (So $\mathcal{F}(n)$ is empty if n is odd. Then it is known that

$$\det A = \Big(\sum_{\pi\in\mathcal{F}(n)} \operatorname{sig}(\pi)\operatorname{wt}(\pi)\Big)^2. \tag{5.7}$$

Here $\operatorname{wt}(\pi) = \prod_{i=1}^{n} a_{i,(i)\pi}$ and $\operatorname{sig}(\pi) = \pm 1$. (The exact definition of $\operatorname{sig}(\pi)$ will not be needed.) The sum here is known as the *Pfaffian* of A. For more information about the Pfaffian, the reader is referred to Godsil (1993, chapter 7), Lovász (1979a), Stembridge (1990), or Northcott (1984).

Suppose now that we are given a graph G, and that we wish to determine whether it has a perfect matching. This can be done as follows. Let $\tilde{A} = a_{ij}$ be a skew-symmetric matrix such that $a_{ij} = 0$ if i and j are not adjacent in G and, moreover, the numbers $\{a_{ij}\colon i < j, ij \in E(G)\}$ are algebraically independent over the rationals. Then from (5.7) we see that $\det\tilde{A}$ is non-zero if and only if G has a perfect matching. This fact, together with Lemma 5.8, was used by Tutte to derive his characterisation of graphs with no perfect matchings.

Instead of choosing the entries of $\tilde{A}$ to be algebraically independent, we can also choose them at random. If $\det\tilde{A} \neq 0$ then G must have a perfect matching. If $\det\tilde{A} = 0$ then we are left uncertain, but by repeating the experiment a number of times we can reduce the uncertainty to any desired level. This strategy was first suggested in Edmonds (1967), for bipartite graphs. For an elegant implementation of this idea and some related background information, see Mulmuley et al. (1987).

6. Eigenvalue methods

In this section our study of adjacency matrices is continued, but now our emphasis will be on their eigenvalues, rather than on walks. We confine ourselves almost entirely to graphs, which means that our adjacency matrices will be symmetric and their eigenvalues real. A great deal of effort has been devoted to the study of the relation between the structure of a graph G and the eigenvalues of $A(G)$. Although this subject has considerable independent interest, we confine ourselves almost entirely to its applications. We begin by introducing two fundamental results from matrix theory, the first of which is a version of the well-known Perron–Frobenius theorems. (See, e.g., Cvetković et al. 1980, Theorem 0.3.)

Theorem 6.1. *Let G be a connected graph. Then the largest eigenvalue ρ of $A(G)$ is simple, and the entries of the corresponding eigenvector are all positive. If λ is any other eigenvalue of $A(G)$ then $\lambda \geqslant -\rho$, with equality holding if and only if G is bipartite: The largest eigenvalue of any proper subgraph of G is less than ρ.*

(The most general, and most natural, form of the Perron–Frobenius theorem is concerned with non-negative matrices; the above version suffices for most of what we need.) If G has maximum degree Δ and largest eigenvalue ρ then $\sqrt{\Delta} \leqslant \rho \leqslant \Delta$. The first inequality holds because the complete bipartite graph $K_{i,\Delta}$ is a subgraph of G and the second because G can be realised as a subgraph of a Δ-regular graph. (This also shows that we can have $\rho = \Delta$ if and only if G is regular.)

Theorem 6.2. *Let u be a vertex in the graph G. Then the eigenvalues of $G \setminus u$ interlace those of G (i.e., between any two eigenvalues of $G \setminus u$ there lies an eigenvalue of G).*

Proof. Assume that G has n vertices and let $A = A(G)$. If U is a subspace of $\mathbb{R}^n$, define $\lambda_U(A)$ to be the minimum value of $x^{\mathrm{T}}Ax$ as x ranges over the unit vectors in U. Denote the kth largest eigenvalue of A by $\lambda_k(A)$. It is known that

$$\lambda_k(A) = \max_{\dim(U)=k} \min_{x \in U} \frac{x^{\mathrm{T}}Ax}{x^{\mathrm{T}}x}.$$

Let S be an $m \times n$ matrix with orthonormal rows, i.e., $SS^{\mathrm{T}} = I_m$. Then we have

$$\lambda_k(SAS^{\mathrm{T}}) = \max_{\dim(U)=k} \lambda_U(SAS^{\mathrm{T}}) = \max_{\dim(U)=k} \lambda_{SU}(A)$$

whence it follows that

$$\lambda_k(SAS^{\mathrm{T}}) \leqslant \lambda_k(A). \tag{6.1}$$

Applying the same argument to $-A$, we further deduce that for $k = 0, \ldots, m$,

$$\lambda_{m-k}(SAS^{\mathrm{T}}) \geqslant \lambda_{n-k}(A). \tag{6.2}$$

If we now choose S to consist of $n-1$ rows of the identity matrix I_n then we obtain the theorem. □

The interlacing property of the eigenvalues of symmetric matrices was first noted in mechanics, arising in the study of the behaviour of a (mechanical) system as new constraints are imposed on its parameters. The proof we have given is based on Haemers (1979). Haemers has used eqs. (6.1) and (6.2) above to obtain a number of interesting results in graph theory and design theory. It is worth noting that there is a connection here to the theory of quotients. Suppose that, in our usual notation, we have $PA = \Phi P$ where P is the characteristic matrix of an equitable partition. Choose Λ to be the non-negative diagonal matrix such that $\Lambda^2 = PP^{\mathrm{T}}$. Then $\Lambda^{-1}PA = \Lambda^{-1}\Phi\Lambda \cdot \Lambda^{-1}P$ and so we may set $S = \Lambda^{-1}P$ and $\Gamma = \Lambda^{-1}\Phi\Lambda$ to obtain $SAS^{\mathrm{T}} = \Gamma$. The rows of S are pairwise orthogonal and thus the inequalities (6.1) and (6.2) follow.

Theorem 6.2 implies that any eigenvalue of G with multiplicity greater than one must also be an eigenvalue of any vertex-deleted subgraph $G\backslash u$. Another consequence is that the least eigenvalue of $G\backslash u$ is bounded below by the least eigenvalue of G. Thus, the class of all graphs with least eigenvalue greater than a fixed number α is closed under the operation of taking subgraphs. The study of these classes turns out to be quite interesting, so we discuss it briefly.

Denote the least eigenvalue of G by $\lambda_{\min}(G)$. Since the eigenvalues of K_2 are -1 and 1, it follows that $\lambda_{\min(G)} \leqslant -1$ for any graph G with at least one edge. The eigenvalues of $K_{1,2}$ are $-\sqrt{2}$, 0 and $\sqrt{2}$, whence we deduce that if G is connected and not complete then $\lambda_{\min(G)} \leqslant -\sqrt{2}$. A more interesting case is the class of graphs with $\lambda_{\min} \geqslant -2$. It can be shown that all line graphs have this property, along with the so-called "generalised line graphs". Considerable effort was devoted to characterising the remaining graphs in this class before Cameron et al. (1976) produced a short, ingenious and elegant solution.

Their work was all the more interesting in that it was based on a connection with the theory of root systems. We outline the way this arises. Let G be a graph with vertex set $V(G) = \{1, \ldots, n\}$ such that $A(G) + 2I$ is a positive semidefinite matrix. There is a matrix X, with linearly independent columns, such that $A(G) + 2I = XX^{\mathrm{T}}$. Let x_i be the ith row of X. Then

$$(x_i, x_j) = \begin{cases} 2, & \text{if } i = j; \\ 1, & \text{if } i \sim j; \\ 0, & \text{otherwise.} \end{cases}$$

Let $\mathscr{L}$ be the lattice formed by the set of all integral combinations of the columns of X. If x is a row of X then the mapping

$$a \mapsto a - (a, x)x$$

fixes $\mathscr{L}$. (Note that this mapping represents reflection in the hyperplane in $\mathbb{R}^m$ perpendicular to x.) From this it follows that the vectors x_i, for i in $V(G)$, are a subset of a *root system*. (For an elementary and pleasant introduction to root systems, see Grove and Benson 1985.)

It would appear that this topic is far from being exhausted. Neumaier (1979) showed that, with finitely many exceptions, the strongly regular graphs with

$\lambda_{\min} = -k$ (for some positive integer k) belong to one of two infinite families. (The strongly regular graphs G with $\lambda_{\min(G)}$ not an integer fall into a third infinite family.) Hoffman (1977) shows that a graph with $\lambda_{\min} \geqslant -1 - \sqrt{2}$ and having "large" valency is a generalised line graph, and consequently has least eigenvalue at least -2. (Here "large" is determined by Ramsey theory, and is thus only technically finite.) This is an intriguing result.

The eigenvalues of a graph also give information about its chromatic number, and related quantities. Let $\lambda_{\max}(G)$ denote the largest eigenvalue of G. We denote the chromatic number of G by $\chi(G)$.

Theorem 6.3 (Hoffman 1970). *The chromatic number of a graph G is bounded below by $1 - \lambda_{\max}(G)/\lambda_{\min}(G)$.*

Proof. Let z be an orthonormal eigenvector of G with eigenvalue $\lambda_{\max}(G)$. Assume that G can be properly coloured with c colours. Such a colouring determines a partition of $V(G)$ with c cells and characteristic matrix P. Let $\tilde{P}$ be the matrix constructed from P by replacing the non-zero entry in column i of P by the corresponding entry of z, and then deleting all zero rows. The rows of $\tilde{P}$ are not orthonormal, but there is a unique non-negative diagonal matrix Δ such that the rows of $S := \Delta\tilde{P}$ are. There is also a vector y such that $y^{\mathrm{T}}S = z^{\mathrm{T}}$. Consequently

$$y^{\mathrm{T}}SAS^{\mathrm{T}}y = z^{\mathrm{T}}Az = \lambda_{\max}(G),$$

which implies that $\lambda_{\max}(G) \leqslant \lambda_{\max}(SAS^{\mathrm{T}})$. On the other hand, since the rows of S are pairwise orthonormal, inequalities (6.1) and (6.2) apply. Thus we deduce that $\lambda_{\max}(G) = \lambda_{\max}(SAS^{\mathrm{T}})$ and accordingly that

$$(c-1)\lambda_{\min}(SAS^{\mathrm{T}}) + \lambda_{\max}(G) = (c-1)\lambda_{\min}(SAS^{\mathrm{T}}) + \lambda_{\max}(SAS^{\mathrm{T}}).$$

By (6.2), the left-hand side is bounded below by $(c-1)\lambda_{\min}(G) + \lambda_{\max}(G)$. The right-hand side is bounded above by $\operatorname{tr} SAS^{\mathrm{T}}$. It is easy to see that the diagonal entries of SAS^{T} are all zero, hence the sum of its eigenvalues is zero. This implies that

$$(c-1)\lambda_{\min}(G) + \lambda_{\max}(G) \leqslant 0$$

and this yields the theorem. □

In deriving Theorem 6.3 we did not use the fact that the non-zero entries of A are all equal to 1; in fact a careful reading will show that we have actually proved that if B is a symmetric matrix such that $(B)_{ij} = 0$ whenever i and j are non-adjacent vertices in G then

$$\chi(G) \geqslant 1 - \lambda_{\max}(B)/\lambda_{\min}(B).$$

If $\lambda_{\min}(B) = -\tau$ then $C := I + \tau^{-1}B$ is a positive semidefinite matrix with diagonal entries equal to 1 and with $(C)_{ij} = 0$ whenever i and j are distinct non-adjacent vertices in G. This leads us to the following.

Corollary 6.4. *Let G be a graph on n vertices and let $\Omega(G)$ be the set of all positive semidefinite matrices C such that $(C)_{ii} = 1$ for all vertices i of G, and $(C)_{ij} = 0$ whenever i and j are distinct non-adjacent vertices. Then*

$$\chi(G) \geqslant \max_{C \in \Omega(G)} \lambda_{\max}(C).$$

The complement of the graph G will be denoted by $\overline{G}$. The quantity

$$\max\{\lambda_{\max}(C) \mid C \in \Omega(\overline{G})\}$$

is usually denoted by $\theta(G)$. Thus Corollary 6.4 asserts that $\chi(G) \geqslant \theta(\overline{G})$. Now suppose that the vertices in the subset S of $V(G)$ induce a complete subgraph of G. Let C_S be the 01-matrix with ij-entry equal to 1 when i and j both lie in C, and equal to zero otherwise. Then $C_S \in \Omega(G)$ and $\lambda_{\max}(C_S) = |S|$. This shows that $\theta(\overline{G}) \geqslant \alpha(\overline{G})$, or equally that $\theta(G) \geqslant \alpha(G)$. (Here $\alpha(G)$ is the maximum number of vertices in an independent set from G.)

The quantity $\theta(G)$ was first introduced in Lovász (1979b), where he established that it provides an upper bound on the "Shannon capacity" of G. We discuss this briefly. If G and H are graphs, let us denote by $G \times H$ their *strong product*. This can be defined as the graph with

$$(A(G) + I) \otimes (A(H) + I) - I$$

as its adjacency matrix. (Thus the vertex set of $G \times H$ is the Cartesian product of $V(G)$ and $V(H)$, and the pairs (u, v) and (u', v') are adjacent if and only if u is equal or adjacent to u' in G and v is equal or adjacent to v' in H. The strong product of n copies of G will be denoted by G^n. It is not hard to show that $\alpha(G \times H) \geqslant \alpha(G)\alpha(H)$ and from this one can deduce that the *Shannon capacity*

$$\Theta(G) := \limsup(\alpha(G^n)^{1/n})$$

exists. The significance of $\theta(G)$ stems from the facts that it is an upper bound for $\alpha(G)$, and that it is multiplicative, i.e., $\theta(G \times H) = \theta(G)\theta(H)$. Together these imply that $\Theta(G) \leqslant \theta(G)$. (For the proof that $\theta(G)$ is multiplicative we refer the reader to Lovász 1979b.) Note that it is not difficult to verify that $\Omega(G \times H)$ contains $\Omega(G) \otimes \Omega(H)$, and this implies that $\theta(G \times H) \geqslant \theta(G)\theta(H)$. It is proved in Grötschel et al. (1981) that $\theta(G)$ can be computed in polynomial time. Lovász found a number of different expressions for $\theta(G)$. One of these is, in a sense, dual to our definition.

Theorem 6.5 (Lovász 1979b). *For any graph G, let $\mathcal{M}(G)$ denote the set of all positive semidefinite matrices such that $\operatorname{tr} B = 1$ and $(B)_{ij} = 0$ if i and j are distinct vertices of G. Then*

$$\theta(G) = \min_{B \in \mathcal{M}(G)} \operatorname{tr}(JB).$$

Using the theory he developed, Lovász was able to deduce the value of $\Theta(G)$ in many new cases. (The smallest of these was C_5, the cycle on five vertices, while $\Theta(C_7)$ is still unknown. This gives some idea of the difficulty of this problem.) Haemers found a simple argument which sometimes provides a better bound on $\Theta(G)$ than $\theta(G)$ does. He observed that, if $\lambda \neq 0$, then the submatrix of $A(G) + \lambda I$ corresponding to an independent set on s vertices is just λI_s. Hence it is nonsingular and so we deduce that

$$\alpha(G) \leqslant \operatorname{rank}(A + \lambda I).$$

From this it can be shown that $\operatorname{rank}(A + \lambda I)$ is an upper bound on $\Theta(G)$. For more information, and examples where this bound is better than $\theta(G)$, see Haemers (1981).

Eigenvalue methods have also been applied to graph factorisation problems. The next example is possibly the best known of these.

Lemma 6.6 (Graham and Pollak 1972). *The edge set of K_n cannot be partitioned into fewer than $n - 1$ complete bipartite subgraphs.*

Proof. Let G be graph on n vertices that is the edge-disjoint union of subgraphs $H_1, \ldots, H_r$. Assume that each of these subgraphs H_i is a spanning subgraph of G consisting of a complete bipartite graph, together with some isolated vertices. We assume without proof the easily established fact that if H is a complete bipartite graph on m vertices then there is an m-dimensional subspace U of $\mathbb{R}^m$ such that the inner product $(u, A(H)u)$ is non-negative for all u in U. (In fact U is spanned by the eigenvectors of $A(H)$ with non-negative eigenvalues.) We say that U is *non-negative* for $A(H)$. It follows that we can associate to each subgraph H_i an $(n-1)$-dimensional subspace of $\mathbb{R}^n$ that is non-negative for $A(H_i)$.

The intersection of the r subspaces $U_1, \ldots, U_r$ has dimension at least $n - r$ and so, if $r \leqslant n - 2$, there is a 2-dimensional subspace U' of $\mathbb{R}^n$ that is non-negative for the $A(G)$. In U' we can find a non-zero vector z orthogonal to the "all-ones" vector j such that $(z, A(G)z) \geqslant 0$. Now suppose that $G = K_n$. Then $A(G) = J - I$ and so, if z is a non-zero vector orthogonal to j, then $(z, A(G)z) = -(z, z) < 0$. This shows that $r > n - 2$. □

The argument just used can be rephrased in terms of real quadratic forms, and in this setting even shorter proofs of Lemma 6.6 can be found. One corollary of the above proof is that a graph on n vertices with exactly m non-negative eigenvalues cannot be expressed as the edge-disjoint union of fewer than $n - m$ complete bipartite graphs. We note another result that can be proved with the method at hand.

Lemma 6.7 (Schwenk 1983, 1987). *The complete graph on 10 vertices cannot be expressed as the edge-disjoint union of three copies of Petersen's graph.*

Proof. Assume that we have

$$J_{10} - I_{10} = A + B + C,$$

where A, B and C are 01-matrices and A and B are both adjacency matrices of copies of Petersen's graph. It is known that the eigenvalues of Petersen's graph are -2, 1 and 3, and that the eigenvalue 1 has multiplicity six. Let T and U be the eigenspaces associated to the eigenvalue 1 of A and B respectively. Since j is an eigenvector with eigenvalue 3 for both A and B, it follows that T and U both lie in the 9-dimensional subspace of $\mathbb{R}^{10}$ formed by the vectors orthogonal to j. Consequently they must have a non-zero common subspace, which we assume is spanned by a vector z. Then $(J - I)z = -z$ and so $Cz = (-3)z$. Thus C has -3 as an eigenvalue, and so cannot be the adjacency matrix of (a copy of) Petersen's graph. □

Note that the matrix C must be the adjacency matrix of a cubic graph and that, by Theorem 6.1, a cubic graph with least eigenvalue equal to -3 is bipartite. Thus the above method is providing more information than is contained in the statement of the lemma, and it also can easily be applied to other situations. It could, for example, be used to study the possibility of partitioning the edges of K_n into three copies of some given strongly regular graph (on n vertices).

Mohar (1992) develops a relation between graph eigenvalues and Hamiltonicity. One consequence of this theory is a proof that the Petersen graph does not contain a Hamilton cycle. There is an amusing direct proof of this using interlacing, which we now describe. Suppose by way of contradiction that there was a Hamilton cycle in the Petersen graph. Then the line graph $L(P)$ of the Petersen graph would contain an induced copy of C_{10} and so, by interlacing, $\theta_i(C_{10}) \leqslant \theta_i(L(P))$ for $i + 1, \ldots, 10$. But in fact $\theta_7(C_{10}) > \theta_7(L(P))$, so the Hamilton cycle cannot exist. (This argument fails to prove that the Coxeter graph has no Hamilton cycle; it would be very interesting to find an extension of this argument which would work for the Coxeter graph.)

Our next topic is the connection between graph eigenvalues and connectivity. For this it is sometimes convenient to use modified forms of adjacency matrices. We discuss them briefly.

If G is a graph on n vertices, let $\Delta = \Delta(G)$ be the $n \times n$ diagonal matrix with Δ_{ii} equal to the valency of the ith vertex of G. The incidence matrix of $B = B(G)$ of G is the 01-matrix with rows indexed by the vertices of G, columns by the edges and with $(B)_{ij}$ equal to 1 if and only if vertex i is in edge j. Then we have

$$BB^{\mathrm{T}} = \Delta(G) + A(G), \qquad B^{\mathrm{T}}B = 2I + A(\mathscr{L}(G)),$$

where $\mathscr{L}(G)$ denotes the line graph of G. (Remark: since $B^{\mathrm{T}}B$ is positive semidefinite, it follows that $\lambda_{\min}(\mathscr{L}(G)) \geqslant -2$, as we mentioned in the discussion following Theorem 6.2.)

An *orientation* of G can be defined to be a function σ on $V \times V$ such that $\sigma(u, v) = -\sigma(v, u)$, and is zero if u and v are not adjacent. If $\sigma(u, v) = 1$ we call

v the *head* and u the *tail* of the edge $\{u, v\}$. The pair (G, σ) is an oriented graph. The incidence matrix B^σ of (G, σ) is defined by

$$(B^\sigma)_{x,e} = \begin{cases} 1, & \text{if } x \text{ is the head of } e; \\ -1, & \text{if } x \text{ is the tail of } e; \\ 0, & \text{otherwise.} \end{cases}$$

The pertinent property of B^σ is that

$$B^\sigma(B^\sigma)^{\mathrm{T}} = \Delta(G) - A(G). \tag{6.3}$$

Much of our notational effort is gone to waste, since the right-hand side of (6.3) is clearly independent of the orientation σ. We do deduce, however, that $\Delta - A$ is a positive semidefinite matrix. The multiplicity of 0 as an eigenvalue of $\Delta - A$ is equal to the dimension of the null-space of B^σ. This in turn is known to equal $n - c$, where c is the number of connected components of G. (One reference for the unproved assertions here is Biggs 1993.) (If G is bipartite then $\Delta - A$ and $\Delta + A$ are similar matrices. I know of no reference for this. However, in this case it is easy enough to find a diagonal matrix Λ, with diagonal entries equal to ± 1, such that $B^\sigma = \Lambda B$. Then $\Lambda(\Delta - A)\Lambda = \Lambda(\Delta + A)\Lambda$ and, since $\Lambda = \Lambda^{-1}$, this proves the claim.)

Let $\lambda_2(G)$ denote the second smallest of the n eigenvalues of $\Delta - A$. From our remarks above we see that $\lambda_2(G) \neq 0$ if and only if G is connected. A study of the relation between λ_2 and connectivity has been made by Fiedler (1973). We observe that if we delete the first row and column from $\Delta - A$ we obtain a matrix, D say, differing from $\Delta(G \backslash 1) - A(G \backslash 1)$ by the addition of some non-negative terms to its diagonal. From this it can be deduced that the ith eigenvalue of D is at least as large as the ith eigenvalue of $\Delta(G \backslash 1) - A(G \backslash 1)$. Since the eigenvalues of this latter matrix interlace those of $\Delta - A$, we conclude that $\lambda_2(G \backslash 1) \leqslant \lambda_2(G)$. This implies, as noted by Fiedler, that $\lambda_2(G)$ is a lower bound on the vertex-connectivity of G. In fact it can be argued that it is more natural here to consider edge-deleted subgraphs, rather than vertex-deleted subgraphs of G. For if $e \in E(G)$ and $H := G \backslash e$ than the difference between $\Delta(G) - A(G)$ and $\Delta(H) - A(H)$ is a matrix with rank one. This implies that the eigenvalues of $G \backslash e$ interlace those of G.

If $X \subseteq V(G)$, let ∂X denote the number of edges of G with one end in X and the other not in X. We have the following.

Lemma 6.8. *Let G be a graph with nvertices and let X be a subset of $V(G)$. Then*

$$|\partial X| \geqslant \lambda_2(G)|X||V \backslash X|/n.$$

Proof. Let j be the vector with all entries equal to 1. Since the rows and columns of $\Delta - A$ all sum to 0, we always have $(\Delta - A)j = 0$. This implies that

$$\lambda_2(G) = \min\{(z, (\Delta - A)z) \mid (z, j) = 0, \|z\| = 1\}.$$

We also have

$$(z, (\Delta - A)z) = \sum_{ij \in E(G)} (z_i - z_j)^2.$$

Now define z by setting z_i equal to α when $i \in X$, and to β otherwise. Choose α and β so that $(z, j) = 0$ and $\|z\| = 1$. Then $(z, (\Delta - A)z) = |\partial X|(\alpha - \beta)^2$. After some calculation we arrive at the statement of the lemma. □

A more general result, using the same basic approach of "guessing" a trial eigenvector z for λ_2, can be found in Alon and Milman (1985, Lemma 2.1). Their work is devoted to a study of "expanders". We will not discuss these further, but instead refer the reader to chapter 32. This subject is perhaps the most important recent application of graph eigenvalues to combinatorics.

Acknowledgement

I wish to thank the combinatorialists at Queen's University (Kingston), in particular Dominique de Caen and Sylvia Monson. They worked through a draft of this chapter in their seminar, and supplied me with a list of the errors they noted.

7. Appendix: Random walks, eigenvalues, and resistance *(L. Lovász)*

The results of sections 5 and 6 concerning the walk generating functions of graphs are closely related to random walks on graphs and to the theory of finite Markov chains, and also to the electrical resistance of the graph. For more on this topic, see Doyle and Snell (1984), and Lovász (1979a, second edition, chapter 11).

Let G be a d-regular connected graph on n vertices with adjacency matrix A. (Most of the results below extend to non-regular graphs, but the formulations are much simpler for regular graphs. We can reduce the general case to this by adding a sufficient number of loops at each vertex; here, a loop adds only 1 to the degree.)

Consider a *random walk on G*: starting at a node v_0, at each step we are at a vertex v_t, and move to each neighbor with probability $1/d$. Let v_t be the random vertex we are at after t steps. Clearly, the sequence of random vertices $(v_t : t = 0, 1, \ldots)$ is a symmetric Markov chain, and $P = d^{-1}A$ is the matrix of transition probabilities. (In fact, every symmetric Markov chain can be viewed as random walk on a graph, if we allow weighted edges. Most facts mentioned below extend quite naturally to all symmetric Markov chains; many extend even to non-symmetric ones.)

Random walks arise in many models in mathematics and physics. For example, consider the shuffling of a deck of cards. Construct a graph whose vertices are all permutations of the deck, and two of them are adjacent if they come by one shuffle move, depending on how we shuffle. Then repeated shuffle moves correspond to a random walk on this graph (see Diaconis 1988). Many models in statistical mechanics can be viewed as a random walk on the set of states.

Random walks have important algorithmic applications. They can be used to reach "obscure" parts of large sets, and also to generate random elements in large and complicated sets, such as the set of lattice points in a convex body, elements of finite groups (see chapter 27) or the set of perfect matchings in a graph (which, in turn, can be used to the asymptotic enumeration of these objects). See Aleliunas et al. (1979), Jerrum and Sinclair (1989), Dyer et al. (1989) for some of these applications.

The probability $p_{ij}^{(t)}$ that, starting at i, we reach j in t steps is the ij-entry of P^t. We define the *probability generating function* for the random walks on G to be

$$P(G,x) := \sum_{t=0}^{\infty} x^t P^t = (I - xP)^{-1}. \tag{7.1}$$

This is of the same form as the walk generating functions studied earlier, and one can apply much of the theory described in the last two sections.

Since P is symmetric, its eigenvalues are real. A trivial eigenvalue of P is 1, with the corresponding eigenvector $(1,\ldots,1)^{\mathrm{T}}$. It follows from the Frobenius–Perron theory that this eigenvalue is unique and that P has spectral radius 1. The value -1 is an eigenvalue of P iff G is bipartite.

Let $1 = \lambda_1 \geqslant \cdots \geqslant \lambda_n$ be the eigenvalues of P (these are just the eigenvalues of A divided by d), and let $v_1,\ldots,v_n$ be corresponding eigenvectors (normed to unit length). Let $v_k = (v_{k1},\ldots,v_{kn})^{\mathrm{T}}$. Clearly we can take $v_{1i} = 1/\sqrt{n}$.

Expressing A in terms of its eigenvectors, we get

$$A = \sum_{k=1}^{n} \lambda_k v_k v_k^{\mathrm{T}}$$

and hence

$$p_{ij}^{(t)} = \sum_{k=1}^{n} \lambda_k^t v_{ki} v_{kj} = \frac{1}{n} + \sum_{k=2}^{n} \lambda_k v_{ki} v_{kj}. \tag{7.2}$$

We shall see how this basic formula can be applied in the analysis of random walks; but first let us introduce some parameters that are significant in the algorithmic applications mentioned above.

(a) The *mean access time* τ_{ij} is the expected number of steps required to reach a vertex j, starting from a vertex i. The sum $\gamma_{ij} = \tau_{ij} + \tau_{ji}$ is called the *mean commute time*.

(b) The *mean cover time* is the expected number of steps to reach every vertex (starting at the vertex for which this is maximum).

(c) The *mixing rate* is a measure of how fast the random walk converges to its limiting distribution. (How long should we shuffle a pack of cards?) This can be defined as follows. If the graph is non-bipartite, then $p_{ij}^{(t)} \to 1/n$ as $t \to \infty$, and the mixing rate is

$$\mu = \limsup_{t\to\infty} \max_{i,j} \left| p_{ij}^{(t)} - \frac{1}{n} \right|^{1/t}.$$

(For a bipartite graph with bipartition $\{V_1, V_2\}$, the distribution of v_t oscillates between "almost uniform on V_1" and "almost uniform on V_2". The results for bipartite graphs are similar, just a bit more complicated to state, so we ignore this case.)

We have to walk about $(\log n)(1-\mu)^{-1}$ steps before the distribution of v_t will be close to uniform. The surprising fact, allowing the algorithmic applications mentioned above, is that this number may be much less than the number of nodes; for an expander, for example, this takes only $O(\log n)$ steps.

An algebraic formula for the mixing rate is easily obtained. Let $\lambda = \max\{|\lambda_2|, |\lambda_n|\}$, then it follows by (7.2) that

$$\left|p_{ij}^{(t)} - \frac{1}{n}\right| < \lambda^t \sum_{k=2}^{n} |v_{ki}v_{kj}| < \lambda^t .$$

So $\mu \leqslant \lambda$; it is not difficult to argue that equality must hold here.

Theorem 7.1. *The mixing rate of a random walk on a non-bipartite graph G is* $\max\{|\lambda_2|, |\lambda_n|\}$.

Lemma 6.8 has established a connection between the second-largest eigenvalue of A (equivalently, of P) and a certain edge-connectivity property of the graph. We define the *conductance* $\Phi = \Phi(G)$ of the graph G as the minimum of $n|\partial X|/(d|X||V \setminus X|)$ over all non-empty sets $X \subset V$. Combining Lemma 6.8 with results of Jerrum and Sinclair (1989) we obtain the following (cf. Alon 1986, Diaconis and Stroock 1991, and also chapter 32, Theorems 3.1 and 3.2).

Theorem 7.2. $\Phi^2/4 \leqslant 1 - \lambda_2 \leqslant \Phi$.

Corollary 7.3.

$$\left|p_{ij}^{(t)} - \frac{1}{n}\right| \leqslant \left(1 - \frac{\Phi^2}{4}\right)^t .$$

The mean access time and the mean commute time can be estimated by elementary means (but, as we shall see later, eigenvalues provide more powerful formulas). We remark first that in a very long random walk, every vertex is visited on the average in every nth step and every edge is traversed in each direction on the average in every $(2m)$th step, where m is the number of edges. (This second assertion remains valid also for random walks over non-regular graphs.) Hence it follows that if we start from node i, and j is an adjacent node, then within $2m$ steps we can expect to pass through the edge ji; hence the mean commute time for two adjacent nodes is bounded by $2m$. It follows that the mean commute time between two nodes at distance r is at most $2mr < n^3$. A similar bound follows for the mean cover time.

Let $q_{ij}^{(t)}$ denote the probability that the random walk starting at i hits vertex j the first time in the tth step. Then we have the following identity by easy case distinction:

$$p_{ij}^{(t)} = \sum_{s=0}^{t} q_{ij}^{(s)} p_{jj}^{(t-s)} .$$

Hence we get for the generating functions $f_{ij}(x) = \sum_{t=0}^{\infty} p_{ij}^{(t)} x^t$ and $g_{ij}(x) = \sum_{t=0}^{\infty} q_{ij}^{(t)} x^t$ that

$$f_{ij}(x) = g_{ij}(x) f_{jj}(x).$$

Now here

$$f_{ij}(x) = \sum_{t=0}^{\infty} \sum_{k=1}^{n} v_{ki} v_{kj} \lambda^t x^t = \sum_{k=1}^{n} \frac{v_{ki} v_{kj}}{1 - \lambda_k x},$$

and so

$$g_{ij}(x) = \sum_{k=1}^{n} \frac{v_{ki} v_{kj}}{1 - \lambda_k x} \Bigg/ \sum_{k=1}^{n} \frac{v_{kj}^2}{1 - \lambda_k x}$$

Now $\tau_{ij} = g'_{ij}(1)$; from this explicit formula we get the following.

Theorem 7.4. *The mean access time is given by*

$$\tau_{ij} = n \sum_{k=2}^{n} \frac{v_{kj}^2 - v_{ki} v_{kj}}{1 - \lambda_k}.$$

Corollary 7.5. *The mean commute time is given by*

$$\gamma_{ij} = n \sum_{k=2}^{n} \frac{(v_{ki} - v_{kj})^2}{1 - \lambda_k}.$$

Since the vectors $u_i = (v_{ik})_{k=1}^n$ are mutually orthogonal unit vectors, we can derive the following bound on the mean commute time between any pair of nodes:

$$\gamma_{ij} = n \sum_{k=2}^{n} \frac{(v_{ki} - v_{kj})^2}{1 - \lambda_k} \leqslant n \frac{1}{1 - \lambda_2} \sum_{k=2}^{n} (v_{ki} - v_{kj})^2$$
$$= n \frac{1}{1 - \lambda_2} (u_i - u_j)^2 = 2n/(1 - \lambda_2).$$

Using Theorem 7.2, we get

$$\gamma_{ij} \leqslant \frac{8n}{\Phi^2},$$

which is better than the elementary bound if, e.g., the graph is an expander: in this case we obtain that $\gamma_{ij} = O(n)$. It also follows from Corollary 7.5 that the mean commute time between any two vertices of any regular graph on n nodes is at least n, so this is best possible for expanders. The best known bound for the mean commute time in a general regular graph is $O(n^2)$, which follows from the analogous bound for the mean cover time below (see Brightwell and Winkler 1990 for the best possible bound in the case of general graphs).

No eigenvalue formula for the mean cover time is known, but a rather good bound follows by elementary probability theory (Matthews 1988).

Proposition 7.6. *The mean cover time of a random walk on a graph with n vertices is at most* $(1+\frac{1}{2}+\cdots+1/n)$ *times the maximum of the mean access times between all pairs of vertices.*

The mean cover time of a regular graph is $O(n^2)$ (Kahn et al. 1989; this issue of *J. Theoret. Probab.* contains many other interesting papers on this problem). This gives a surprisingly narrow range for cover times. It is conjectured that the graph with smallest cover time is the complete graph (whose cover time is $\approx n \log n$). This was recently proved in the asymptotic sense by Feige (1993).

There is an interesting connection between random walks on graphs and electrical networks. We may consider a graph G on n vertices as an electrical network, every edge corresponding to unit resistance. The network has some resistance R_{ij} between any pair of vertices. A whole book has been written on this connection (Doyle and Snell 1984, see also Spitzer 1976); here we only formulate one surprising identity (Nash-Williams 1959, Chandra et al. 1989):

Theorem 7.7. *The mean commute time between vertices i and j is* ndR_{ij}.

The proof (which is only sketched) is connected to yet another interesting notion. We call a function $\phi : V(G) \to \mathbb{R}$ *harmonic* with poles s and t if

$$\sum_{i\in N(j)} \phi(i) = d\phi(j)$$

for every $j \neq s, t$. It is easy to see that if we normalize so that $\phi(s) = 1$ and $\phi(t) = 0$, then the harmonic function with given poles is uniquely determined.

There are (at least) two rather natural ways to construct such harmonic functions.

(1) Consider the graph as an electrical network as above. Give voltage 1 to s and voltage 0 to t. Then the voltage $\phi(i)$ of vertex i defines a harmonic function.

(2) Let $\phi(i)$ denote the probability that a random walk starting at i hits s before it hits t. It is trivial that this defines a harmonic function.

Now the resistance R_{st} is 1/(total current) $= 1/\sum_{i\in N(t)} \phi(i)$. On the other hand, consider a very long random walk, with K steps, say. This hits t about K/n times. Call a hit *interesting* if after it the random walk hits s before it hits t again. Between two interesting hits, the average number of steps is γ_{st}. Now the probability that a given hit is interesting is $1/d\sum_{i\in N(t)} \phi(i)$, by interpretation (2) of the harmonic function. Hence the number of interesting hits is about $1/d\sum_{i\in N(t)} \phi(i)(K/n)$, and so the average number of steps between them is $nd/(\sum_{i\in N(t)} \phi(i)) = ndR_{st}$.

References

Aigner, M.
[1979] *Combinatorial Theory* (Springer, Berlin).

Aleliunas, B., R.M. Karp, R.J. Lipton, L. Lovász and C.W. Rackoff
[1979] Random walks, universal travelling sequences, and the complexity of maze problems, in: *Proc. 20th*

Annu. Symp. on Foundations of Computer Science (IEEE Computer Society Press, New York) pp. 218–223.

Alon, N.
[1985] An extremal property for sets with applications to graph theory, *J. Combin. Theory A* **40**, 82–89.
[1986] Eigenvalues and expanders, *Combinatorica* **6**, 83–96.

Alon, N., and V.D. Milman
[1985] λ_1-Isoperimetric inequalities for graphs and superconcentrators, *J. Combin. Theory B* **38**, 73–88.

Anstee, R.P.
[1985] General forbidden configuration theorems, *J. Combin. Theory A* **40**, 108–124.

Bannai, E.
[1977] On tight designs, *Quart. J. Math. (Oxford)* **28**, 433–448.

Biggs, N.L.
[1993] *Algebraic Graph Theory*, 2nd Ed. (Cambridge University Press, Cambridge).

Bollobás, B.
[1965] On generalized graphs, *Acta Math. Acad. Sci. Hungar.* **16**, 447–452.

Bondy, J.A., and R.L. Hemminger
[1977] Graph reconstruction – a survey, *J. Graph. Theory* **1**, 227–268.

Brightwell, G., and P. Winkler
[1990] Maximum hitting time for random walks on graphs, *J. Random Structures & Algorithms* **1**, 263–276.

Cameron, P.J.
[1976] Transitivity of permutation groups on unordered sets, *Math. Z.* **148**, 127–139.

Cameron, P.J., and J.H. Van Lint
[1991] *Designs, Graphs, Codes and their Links* (Cambridge University Press, Cambridge).

Cameron, P.J., J.-M. Goethals, J.J. Seidel and E.E. Shult
[1976] Line graphs, root systems and elliptic geometry, *J. Algebra* **43**, 305–327.

Chandra, A.K., P. Rahavan, W.L. Ruzzo, R. Smolensky and P. Tiwari
[1989] The electrical resistance of a graph captures its commute and cover times, in: *Proc. 21st ACM STOC* (ACM, New York) pp. 574–586.

Cvetković, D.M., M. Doob and H. Sachs
[1980] *Spectra of Graphs* (Academic Press, New York).

De Caen, D., and D.A. Gregory
[1985] Partitions of the edge-set of a multigraph by complete graphs, *Congress. Numerantium* **47**, 255–263.

Delsarte, Ph.
[1973] The associations schemes of coding theory, *Philips Research Reports Suppl.* No. 10.

Delsarte, Ph., J.-M. Goethals and J.J. Seidel
[1977] Spherical codes and designs, *Geom. Dedicata* **6**, 363–388.

Dembowski, P.
[1968] *Finite Geometries* (Springer, Berlin).

Diaconis, P.
[1988] *Group Representations in Probability and Statistics* (Institute for Mathematical Statistics, Hayward, CA).

Diaconis, P., and D. Stroock
[1991] Geometric bounds for eigenvalues of Markov chains, *Ann. Appl. Probab.* **1**, 36–61.

Dowling, T.A., and R.M. Wilson
[1975] Whitney number inequalities for geometric lattices, *Proc. Amer. Math. Soc.* 47, 504–512.

Doyle, P.G., and J.L. Snell
[1984] *Random Walks and Electrical Networks* (Mathematical Association of America, Washington, DC).

Dyer, M., A. Frieze and R. Kannan
[1989] A random polynomial time algorithm for approximating the volume of convex bodies, in: *Proc. 21st Annu. ACM Symp. on Theory of Computing* (ACM, New York) pp. 375–381.

Edmonds, J.R.
[1967] Systems of distinct representatives and linear algebra, *J. Res. Nat. Bur. Standards B* **71**, 241–247.

Evans, D.M.

[1986] Homogeneous geometries, *Proc. London Math. Soc. (3)* **52**, 305–327.

Farrell, E.J.

[1979] An introduction to matching polynomials, *J. Combin. Theory B* **27**, 75–86.

Feige, U.

[1993] *A Tight Lower Bound on the Cover Time for Random Walks on Graphs*, Tech. Report CS93-19 (Weizman Institute).

Fiedler, M.

[1973] Algebraic connectivity of graphs, *Czech Math. J.* **23**, 298–305.

Fisher, R.A.

[1940] An examination of the possible different solutions of a problem in incomplete blocks, *Ann. Eugenics (London)* **10**, 52–75.

Foody, W., and A. Hedayat

[1977] On theory and application of BIB designs with repeated blocks, *Ann. Statist.* **5**, 932–945.

Frankl, P., and J. Pach

[1983] On the number of sets in a null t-design, *European J. Combin.* **4**, 21–23.

Frankl, P., and R.M. Wilson

[1981] Intersection theorems with geometric consequences, *Combinatorica* **1**, 357–368.

Godsil, C.D.

[1981a] Matching behaviour is asymptotically normal, *Combinatorica* **4**, 369–376.

[1981b] Matchings and walks in graphs, *J. Graph. Theory* **5**, 285–297.

[1988/89] Polynomial spaces, *Discrete Math.* **73**, 71–88.

[1993] *Algebraic Combinatorics* (Chapman and Hall, New York).

Godsil, C.D., and I. Gutman

[1981] On the theory of the matching polynomial, *J. Graph Theory* **5**, 137–144.

Godsil, C.D., and B.D. McKay

[1980] Feasibility conditions for the existence of walk-regular graphs, *Linear Algebra Appl.* **30**, 51–61.

[1981] Spectral conditions for reconstructibility of a graph, *J. Combin. Theory B* **30**, 285–289.

Godsil, C.D., I. Krasikov and Y. Roddity

[1987] Reconstructing graphs from their s-edge deleted subgraphs, *J. Combin. Theory B* **43**, 360–363.

Gottlieb, D.H.

[1966] A certain class of incidence matrices, *Proc. Amer. Math. Soc.* **17**, 1233–1237.

Graham, R., and H.O. Pollak

[1972] On embedding graphs in squashed cubes, in: *Graph Theory and Applications, Lecture Notes in Mathematics*, Vol. 303, eds. Y. Alavi, D.R. Lick and A.T. White (Springer, Berlin) pp. 99–110.

Graham, R., S.-Y. Li and W.-C. Li

[1980] On the structure of t-designs, *SIAM J. Algebraic Discrete Methods* **1**, 8–14.

Grötschel, M., L. Lovász and A. Schrijver

[1981] The ellipsoid method and its consequences in combinatorial optimization, *Combinatorica* **1**, 169–197.

Grove, L.C., and C.T. Benson

[1985] *Finite Reflection Groups*, 2nd Ed. (Springer, New York).

Haemers, W.H.

[1979] Eigenvalue methods, in: *Packing and Covering in Combinatorics*, ed. A. Schrijver (Mathematisch Centrum, Amsterdam) pp. 15–38.

[1981] An upper bound for the Shannon capacity of a graph, in: *Algebraic Methods in Graph Theory*, Vol. 1, eds. L. Lovász and V.T. Sós (North-Holland, Amsterdam) pp. 267–272.

Hajnal, A., W. Maass and G. Turán

[1988] On the communication complexity of graph properties, in: *Proc. 20th Annu. ACM Symp. on Theory of Computing* (ACM Press, New York) pp. 186–191.

Heilmann, O.J., and E.H. Lieb

[1972] Theory of monomer–dimer systems, *Comm. Math. Phys.* **25**, 190–232.

Hoffman, A.J.

[1970] On eigenvalues and colorings of graphs, in: *Graph Theory and its Applications*, ed. B. Harris (Academic Press, New York) pp. 79–91.

[1977] On graphs whose least eigenvalue exceeds $-1-\sqrt{2}$, *Linear Algebra Appl.* **16**, 153–165.

Hoffman, A.J., and R.R. Singleton

[1960] On Moore graphs with diameters 2 and 3, *IBM J. Res. Develop.* **4**, 497–504.

Hughes, D., and F. Piper

[1973] *Projective Planes* (Springer, Berlin).

Jacobi, C.G.J.

[1833] De binis quibuslibet functionibus homogeneis secundi ordinis per substitutiones lineares in alias binas transformandis, quae solis quadratis variabilium constant; una cum variis theorematis de transformatione et determinatione integralium multiplicium, *Crelle's J.* **12**, 1–69; *Werke*, Vol. III, pp. 191–268.

James, G.D.

[1978] *The Representation Theory of Symmetric Groups, Lecture Notes in Mathematics*, Vol. **682** (Springer, Berlin).

Jerrum, M., and A. Sinclair

[1989] Approximating the permanent, *SIAM J. Comput.* **18**, 1149–1178.

Kahn, J., and G. Kalai

[1992] *A Counterexample to Borsuk's Conjecture*, RUTCOR Research Report No. 42–92.

Kahn, J.D., N. Linial, N. Nisan and M.E. Saks

[1989] On the cover time of random walks on graphs, *J. Theory Probab.* **2**, 121–128.

Kantor, W.

[1972] On incidence matrices of finite projective and affine spaces, *Math. Z.* **124**, 315–318.

Koornwinder, T.H.

[1976] A note on the absolute bound for systems of lines, *Indag. Math.* **38**, 152–153.

Kung, J.P.S.

[1985] Matchings and Radon transforms in lattices I. Consistent lattices, *Order* **2**, 105–112.

[1987] Matchings and Radon transforms in lattices II. Concordant sets, *Math. Proc. Cambridge Philos. Soc.* **101**, 221–231.

Lindström, B.

[1969] Determinants on semilattices, *Proc. Amer Math. Soc.* **20**, 207–208.

Lovász, L.

[1972] A note on the line reconstruction problem, *J. Combin. Theory B* **13**, 309–310.

[1976] Chromatic number of hypergraphs and linear algebra, *Studia Sci. Math. Hungar.* **11**, 113–114.

[1977] Flats in matroids and geometric graphs, in: *Combinatorial Surveys*, ed. P.J. Cameron (Academic Press, New York) pp. 45–86.

[1979a] *Combinatorial Problems and Exercises* (North-Holland, Amsterdam). 2nd Edition: 1993.

[1979b] On the Shannon capacity of a graph, *IEEE Trans. Inform. Theory* **IT-25**, 1–7.

[1979c] Topological and algebraic methods in graph theory, in: *Graph Theory and Related Topics*, eds. J.A. Bondy and U.S.R. Murty (Academic Press, New York) p. 1–14.

Lovász, L., and A. Schrijver

[1990] Cones of matrices and 0-1 optimization, *SIAM J. Optim.* **1**, 166–190.

MacWilliams, F.J., and N.J.A. Sloane

[1978] *The Theory of Error-Correcting Codes* (North-Holland, Amsterdam).

Majindar, K.N.

[1962] On the parameters and intersections of blocks of balanced incomplete block designs, *Ann. Math. Statist.* 33, 1200–1205.

Matthews, P.

[1988] Covering problems for Brownian motion on spheres, *Ann. Probab.* **16**, 189–199.

McKay, B.D.

[1979] Transitive graphs with fewer than 20 vertices, *Math. Comp.* **33**, 1101–1121.

Mohar, B.R.

[1992] Domain monotonicity theorem for graphs and Hamiltonicity, *Discrete Appl. Math.* **36**, 169–177.

Müller, V.

[1977] The edge reconstruction conjecture is true for graphs with more than $n \log_2 n$ vertices, *J. Combin. Theory B* **22**, 281–283.

Mulmuley, K., U.V. Vazirani and V.V. Vazirani

[1987] Matching is as easy as matrix inversion, *Combinatorica* **7**, 105–113.

Nash-Williams, C.St.J.A.

[1959] Random walk and electric currents in networks, *Proc. Cambridge Philos. Soc.* **55**, 181–194.

Neumaier, A.

[1979] Strongly regular graphs with least eigenvalue $-m$, *Arch. Math.* **33**, 392–400.

Northcott, D.G.

[1984] *Multilinear Algebra* (Cambridge University Press, Cambridge).

Ray-Chaudhuri, D.K., and R.M. Wilson

[1975] On t-designs, *Osaka J. Math.* **12**, 737–744.

Riordan, J.

[1958] *An Introduction to Combinatorial Analysis* (Wiley, New York).

Schwenk, A.J.

[1983] Advanced problem No. 6434, *Amer. Math. Monthly* **90**, 403.

[1987] Solution to advanced problem No. 6434, *Amer. Math. Monthly* **94**, 885.

Seymour, P.

[1974] On 2-colourings of hypergraphs, *Quart. J. Math. Oxford* **25**, 303–312.

Spitzer, F.

[1976] *Principles of Random Walk* (Springer, New York).

Stanley, R.P.

[1980] Weyl groups, the hard Lefschetz theorem and the Sperner property, *SIAM J. Algebraic Discrete Methods* **1**, 168–184.

[1982] Some aspects of groups acting on finite posets, *J. Combin. Theory A* **32**, 132–161.

[1985] Quotients of Peck posets, *Order* **1**, 29–34.

Stembridge, J.R.

[1990] Nonintersecting paths, Pfaffians, and plane partitions, *Adv. in Math.* **83**, 96–131.

Tutte, W.T.

[1947] The factorisation of linear graphs, *J. London Math. Soc.* **22**, 107–111.

[1979] All the king's horses, in: *Graph Theory and Related Topics*, eds. J.A. Bondy and U.S.R. Murty (Academic Press, New York) pp. 15–33.

Valiant, L.J.

[1979] The complexity of computing the permanent, *Theor. Comput. Sci.* **8**, 189–201.

Wilf, H.S.

[1968] Hadamard determinants, Möbius functions and the chromatic number of a graph, *Bull. Amer. Math. Soc.* **74**, 960–964.

Wilson, R.M.

[1973] The necessary conditions for t-designs are sufficient for something, *Utilitas Math.* **4**, 207–215.

[1990] A diagonal form for the incidence matrices of t-subsets vs. k-subsets, *European J. Combin.* **11**, 609–615.

Yuan, H.

[1982] An eigenvector condition for reconstructibility, *J. Combin. Theory B* **32**, 353–354.

CHAPTER 32

Tools from Higher Algebra

Noga ALON

Department of Mathematics, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Ramat-Aviv, Tel Aviv 69978, Israel

Contents

HANDBOOK OF COMBINATORICS
Edited by R. Graham, M. Grötschel and L. Lovász

1. Introduction

Tools from many areas of mathematics are standard in certain branches of combinatorics and are described in detail in some of the chapters of this Handbook. Examples are the use of linear and multilinear algebra in the theory of designs and in extremal set theory, the use of finite groups in coding theory, application of representation theory of the symmetric group for deriving combinatorial identities, and application of probability theory for obtaining asymptotic existence proofs of combinatorial structures; the use of convexity and linear programming in combinatorial optimization, and the use of topological methods in the study of posets, convex polytopes and various extremal problems.

The objective of this chapter is to survey some sporadic results from several areas of mathematics which were used successfully in solving certain combinatorial problems. It is believed that these results will soon be integrated into the mathematical machinery commonly used in combinatorics. We fully realize the arbitrariness of any such selection and do not claim that these are the most important examples that could be listed. We have, however, no doubt that they merit to be mentioned in this chapter.

The combinatorial applications described here apply various tools from several areas of mathematics. It is natural to wonder whether the use of all these powerful tools is necessary. After all, it is reasonable to believe that combinatorial statements can be proved using combinatorial arguments. Pure combinatorial proofs are desirable, since they might shed more light on the corresponding problems.

No such combinatorial proofs are known for any of the main results discussed in this chapter. It would be nice to try and obtain such proofs.

One of the major consumers of powerful mathematical tools in combinatorics is the area of explicit constructions. The existence of many combinatorial structures with certain properties can be established using the probabilistic method. It is natural to ask for an explicit description of such a structure. Such a construction is particularly valuable when the required structure is needed for solving a certain algorithmic problem. In sections 2-4 we describe several mathematical tools used for explicit constructions. These include the use of group theory for constructing graphs without short cycles, the use of the theory of representations of Lie groups for constructing expanders, and the application of certain results from analytic number theory for constructing pseudo random tournaments. We note that an exact definition of the notion “explicit” (or “uniform”) construction can be given, but we prefer its intuitive meaning here. In sections 5-9 we survey some combinatorial applications of results from other mathematical areas, including real and complex algebraic geometry, algebraic and analytic number theory and hyperbolic geometry.

Most of the mathematical background for the material in this chapter can be found in the books and surveys given in the following list: section 2: Magnus et al. (1966); section 3: Newman (1972); section 4: Schmidt (1976); section 5: Hocking and Young (1961); section 6: Borevich and Shafarevich (1966); section 7: Redei (1973), Van der Waerden (1931); section 8: Coxeter (1956); section 9: Stanley (1983).

2. Group theory and graphs with large girth

The *girth* of a graph G is the length of the shortest cycle in G. If $G=(V,E)$ is a d-regular graph with n nodes and girth $g>2k$, then

$$d\cdot\left(1+(d-1)+\cdots+(d-1)^{k-1}\right)\leqslant n\ ,$$

since the left-hand side of the last inequality is precisely the number of nodes within distance k from a given node v of G. Therefore,

$$g\leqslant 2+2\log n/\log(d-1)\ .$$

Thus, for any fixed $d\geqslant 3$, the girth of a family of d-regular graphs can grow at most at the rate of the logarithm of the number of nodes. Erdős, Sachs, Sauer and Walther (cf. Bollobás (1978), pp. 103-110) proved the existence of d-regular graphs with girth g and n nodes, where

$$g\geqslant\log n/\log(d-1)\ . \tag{2.1}$$

Although their proof does supply a polynomial time algorithm for constructing such graphs, their graphs are not really explicit in the sense that it is not clear how to decide efficiently if two vertices of such a graph are adjacent, given their names.

It seems more difficult to construct explicitly for some fixed $d\geqslant 3$, a family of d-regular graphs whose girth grows at the rate of the logarithm of the number of nodes. Such a construction was first given by Margulis (1982) who used Cayley graphs of factor groups of free subgroups of the modular group. His construction, together with some related results, is outlined below.

2.1. Cayley graphs

Let H be a finite group with a generating set δ satisfying $\delta=\delta^{-1}$, $1\notin\delta$. The *Cayley graph* $G=G(H,\delta)$ is a graph whose nodes are the elements of H in which u and v are adjacent iff $u=sv$ for some $s\in\delta$. Clearly, G is $|\delta|$ regular and a cycle in G corresponds to a reduced word in the generators which represents the identity of H. Cayley graphs are fairly obvious candidates for regular graphs with large girth, since it is not too difficult to see that for every d and g there exists a d-regular Cayley graph with girth at least g. This is equivalent to the group theoretical property of residual finiteness and is proved as follows (see, e.g., Biggs 1989).

Let T be a finite d-regular tree of radius r with center w, whose edges are properly d-colored. Define d permutations $\pi_1,\ldots,\pi_d$ on the nodes of T by putting $\pi_i(u)=v$ if $\{u,v\}$ is an edge of T colored i, and $\pi_i(u)=u$ if u is a leaf of T and the color i is not represented at u. Clearly $\pi_1,\ldots,\pi_d$ are involutions and they generate a group of permutations H. Put $\delta=\{\pi_1,\ldots,\pi_d\}$ and consider the Cayley graph $G=G(H,\delta)$. Consider the effect of a reduced word in the π_i on the central node w. Initially, each element of the first r elements of the word moves w one step towards the boundary. To return the image of w to itself another $r+1$ elements are required. Hence, the girth of G is at least $2r+1$. We note that a more careful analysis will show that the girth of G is, in fact, at least $4r+2$.

The last construction is explicit but gives a much weaker lower bound for g than the one given in (2.1). More efficient solutions can be obtained using familiar groups.

2.2. *The construction of Margulis*

For a commutative ring K with identity, let $\mathrm{S}\ell(2,K)$ denote the group of all two-by-two matrices over K with determinant 1. Consider the integral matrices

$$A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix}$$

and put $\delta = \{A, B, A^{-1}, B^{-1}\}$. For a prime p, let f_p be the natural homomorphism of $\mathrm{S}\ell(2,\mathbb{Z})$ onto $\mathrm{S}\ell(2,\mathbb{Z}_p)$ given by $f_p(a_{ij}) = (a_{ij} \,(\mathrm{mod}\, p))$. Put $A_p = f_p(A)$, $B_p = f_p(B)$ and let G_p be the Cayley graph $G(\mathrm{S}\ell(2,\mathbb{Z}_p), f_p(\delta))$.

Theorem 2.1 (Margulis 1982). *G_p has $n_p = p(p^2-1)$ nodes and is 4-regular. Its girth g_p is at least $2\log_\alpha(p/2) - 1$, where $\alpha = 1 + \sqrt{2}$. Hence $g_p > 0.83 \log n_p / \log 3 - 3$.*

Proof. The first statement is trivial. To bound $g = g_p$, we estimate k, defined as the largest integer such that any two distinct paths in G_p of lengths $< k$ starting at $I = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$ end at different vertices. Clearly $g \geqslant 2k - 1$. Given two such paths $P = (p_0, p_1, \ldots, p_r)$ and $Q = (q_0, q_1, \ldots, q_t)$, starting at $p_0 = q_0 = I$ and ending at $p_r = q_t$, let $V = (v_1, \ldots, v_r)$ and $W = (w_1, \ldots, w_t)$ be the corresponding reduced words over δ_p. Clearly $v_1 \cdots v_r = w_1 \cdots w_t$. Define $\bar{v}_i$ to be A if $v_i = A_p$, B if $v_i = B_p$, A^{-1} if $v_i = A_p^{-1}$, and B^{-1} if $v_i = B_p^{-1}$, and define $\bar{w}_i$ analogously. The crucial fact (cf., e.g., Magnus et al. 1966) is that $\{A, B\}$ generate a free subgroup of $\mathrm{S}\ell(2,\mathbb{Z})$. Thus $\bar{v}_1 \cdots \bar{v}_r \neq \bar{w}_1 \cdots \bar{w}_t$ and since $f_p(\bar{v}_1 \cdots \bar{v}_r) = f_p(\bar{w}_1 \cdots \bar{w}_t)$ we conclude that all elements of the non-zero matrix $\bar{v}_1 \cdots \bar{v}_r - \bar{w}_1 \cdots \bar{w}_t$ are divisible by p and hence its norm is at least p. Here the norm $\|L\|$ of a matrix L is $\sup_{x \neq 0} \|Lx\| / \|x\|$. This implies that $\max(\|\bar{v}_1 \cdots \bar{v}_r\|, \|\bar{w}_1 \cdots \bar{w}_t\|) \geqslant p/2$, and since the norms of A, B, A^{-1}, B^{-1} are all $\alpha = 1 + \sqrt{2}$, as is easily checked, we conclude that $\alpha^k \geqslant \alpha^{\max(r,s)} \geqslant p/2$ and $k \geqslant 2\log_\alpha(p/2)$, as needed. □

2.3. *Other constructions*

Modifying the methods of Margulis, Imrich (1984) constructed, for every integer $d \geqslant 3$, infinitely many d-regular Cayley graphs G whose girth $g(G)$ and number of nodes $n(G)$ satisfy

$$g(G) > 0.48 \log n(G) / \log(d-1) - 2 .$$

For $d = 3$ he obtained

$$g(G) > 0.9602 \log n(G) / \log 2 - 5$$

which is very marginally worse than the bound given in (2.1), produced by non-explicit methods.

The best estimate, for infinitely many degrees, was finally obtained by explicit constructions. Using certain algebraic computations in an appropriate algebra of quaternions, Weiss (1984), showed that the members of a certain family of bipartite cubic graphs, explicitly constructed by Biggs and Hoare (1983), have very large girth. The order n and the girth g of each of these graphs satisfy

$$g \geqslant \tfrac{4}{3}\log n/\log 2 - 4 .$$

More generally, Margulis (1984) and, independently, Lubotzky et al. (1986, 1988), constructed, for any prime $p \equiv 1 \pmod 4$, a family of $d = (p+1)$-regular graphs G with

$$g(G) \geqslant \tfrac{4}{3}\log n(G)/\log(d-1) - \log 4/\log(d-1) .$$

Their graphs are Cayley graphs of factor groups of the modular groups, and they have several other interesting properties. These properties are summarized in Theorem 3.5 in the next section. Both constructions improve the bound given in (2.1) and supply one of the rare examples in which an explicit construction improves a non-explicit one.

3. Expanders and superconcentrators

One of the best examples of the use of powerful mathematical tools for explicit constructions is the construction of expanders. For our purposes here, we call a graph $G = (V, E)$ an *(n, d, c)-expander* if it has n nodes, the maximum degree of a node is d, and for every set of nodes $W \subseteq V$ of cardinality $|W| \leqslant n/2$, the inequality $|N(W)| \geqslant c \cdot |W|$ holds, where $N(W)$ denotes the set of all nodes in $V \backslash W$ adjacent to some node in W. We note that the common definition of an expander is slightly different (see, e.g., Gabber and Galil 1981), but the difference is not essential. A family of *linear expanders* of *density* d and *expansion* c is a set $\{G_i\}_{i=1}^{\infty}$ where G_i is an (n_i, d, c)-expander, $n_i \to \infty$ and $n_{i+1}/n_i \to 1$ as $i \to \infty$.

Such a family is the main component of the parallel sorting network of Ajtai et al. (1983), and in the construction of certain fault tolerant linear arrays. It also forms the basic building block used in the construction of graphs with special connectivity properties and small number of edges (see, e.g., Chung 1978).

An example of a graph of this type is an *n-superconcentrator* which is a directed acyclic graph with n inputs and n outputs such that for every $1 \leqslant r \leqslant n$ and every two sets A of r inputs and B of r outputs there are r vertex disjoint paths from the vertices of A to the vertices of B. A family of linear superconcentrators of *density* d is a set $\{G_n\}_{n=1}^{\infty}$, where G_n is an n-superconcentrator with $\leqslant (d + o(1))n$ edges. Superconcentrators, which are the subject of an extensive literature, are relevant to computer science in several ways. They have been used in the construction of graphs that are hard to pebble (see, e.g., Paul et al. 1977), in the study of lower bounds (Valiant 1976), and in the establishment of time space tradeoffs for computing various functions (see, e.g., Tompa 1980).

It is not too difficult to prove the existence of a family of linear expanders (and hence a family of linear superconcentrators) using probabilistic arguments. This was first done by Pinsker (1973) (see also Pippenger 1977 and Chung 1978). However, for applications, an explicit construction is desirable. Such a construction is much more difficult and was first given in the elegant paper of Margulis (1973) who used, surprisingly, some results of Kazhdan on group representations, to construct explicitly a family of linear expanders of density 5 and expansion c, for some $c > 0$. An outline of his method is given below. However, Margulis was not able to bound c strictly away from 0. Gabber and Galil (1981) modified Margulis' construction and were able to give, using Fourier analysis, an effective estimate for c. Better expanders were found later, by several authors, using various methods that are discussed briefly at the end of this section.

3.1. Eigenvalues and expanders

There is a close correspondence between the expansion properties of a graph and the eigenvalues of a certain matrix associated with it. Specifically, let $G = (V,E)$ be a graph and let $A_G = (a_{uv})_{u,v \in V}$ be its adjacency matrix given by $a_{uv} = 1$ if $uv \in E$ and $a_{uv} = 0$ otherwise. Put $Q_G = \operatorname{diag}(\deg(v))_{v \in V} - A_G$, where $\deg(v)$ is the degree of the node $v \in V$, and let $\lambda(G)$ be the second smallest eigenvalue of Q_G. The following simple result is proved in Alon and Milman (1984, 1985). The proof uses elementary linear algebra (Rayleigh's principle). Similar results appear in Tanner (1984), Jimbo and Maruoka (1985) and Buck (1986).

Theorem 3.1. *If G is a graph with n nodes, maximum degree d and $\lambda = \lambda(G)$, then G is an (n,d,c)-expander, where $c = 2\lambda/(d+2\lambda)$.*

Therefore, if $\lambda(G)$ is large then G is a good expander. The converse is also true, though less obvious, and is given in the following result which is in some sense the discrete analogue of a theorem of Cheeger on Riemannian manifolds.

Theorem 3.2 (Alon 1986a). *If G is an (n,d,c)-expander, then $\lambda(G) \geqslant c^2/(4+2c^2)$.*

These two theorems supply an efficient algorithm to approximate the expanding properties of a graph and show that it is enough to estimate $\lambda(G)$ in order to get bounds on the expansion coefficient of G.

3.2. Constructing expanders using Kazhdan's property (T)

Definition. A discrete group H has property (T) if for every set S of generators of H there exists an $\varepsilon > 0$ such that for every unitary representation π of H in $V = V_\pi$ that does not contain the trivial representation, and for every unit vector $y \in V$, there exists an $s \in S$ such that $|(\pi(s)y, y)| < 1 - \varepsilon$.

Kazhdan (1967) defined property (T) for the more general class of locally compact groups. For our purposes the definition for discrete groups suffices.

Margulis, in a beautiful paper, used some of Kazhdan's results on property (T) to construct a family of linear expanders. A somewhat simpler proof for the expansion properties of graphs constructed by a slightly more general construction is given in Alon and Milman (1985). We outline this construction below. Recall the definition, given in section 2, of a Cayley graph $G = G(H, \delta)$, where H is a finite group and δ is a set of generators of H, $\delta = \delta^{-1}$, $1 \notin \delta$.

For $n \geqslant 3$, let $\mathrm{S}\ell(n, \mathbb{Z})$ denote the group of all n by n matrices over the integers $\mathbb{Z}$ with determinant 1. There is a well-known explicit set B_n of two generators of $\mathrm{S}\ell(n, \mathbb{Z})$ (see, e.g., Newman 1972). Put $S_n = B_n \cup B_n^{-1}$ ($|S_n| = 4$). Let $\mathrm{SL}(n, \mathbb{Z}_i)$ be the group of all n by n matrices over the ring of integers modulo i with determinant 1, and let $\phi_i^{(n)} : \mathrm{SL}(n, \mathbb{Z}) \to \mathrm{SL}(n, \mathbb{Z}_i)$ be the group homomorphism defined by $\phi_i^{(n)}((a_{rs})) = (a_{rs} \,(\mathrm{mod}\, i))$. Also define $G_i^{(n)} = G(\mathrm{S}\ell(n, \mathbb{Z}_i), \phi_i^{(n)}(S_n))$.

Theorem 3.3 (Kazhdan 1967). *For each $n \geqslant 3$, $\mathrm{S}\ell(n, \mathbb{Z})$ has property (T).*

It is not too difficult to check that the adjacency matrix $A_i^{(n)}$ of $G_i^{(n)}$ is $\sum_{s \in S_n} \pi \circ \phi_i^{(n)}(s)$, where π is the left regular representation of $\mathrm{S}\ell(n, \mathbb{Z}_i)$. By Rayleigh's principle, $\lambda(G_i^{(n)})$ is precisely the minimum of $|S_n| - (A_i^{(n)} y, y)$, where y ranges over all unit vectors in W which is the space of all vectors whose coordinates sum is zero. Combining these two facts with Theorem 3.4 and the fact that $\pi \circ \phi_i^{(n)}$ is a unitary representation of $\mathrm{S}\ell(n, \mathbb{Z})$ in W that does not contain the trivial representation, we conclude that for every fixed $n \geqslant 3$ there is an $\varepsilon > 0$ such that $\lambda(G_i^{(n)}) \geqslant \varepsilon$ for every i. Hence $\{G_i^{(n)}\}_{i=2}^{\infty}$ is a family of linear expanders of density 4.

3.3. Improved constructions

Various authors have modified and improved Margulis' first construction. Angluin (1979) showed how to construct a family of linear expanders of density 3. Gabber and Galil were the first to obtain a family of linear expanders with an effective estimate on their expansion coefficient. This enabled them to construct superconcentrator of density 271.8. Other constructions appeared in Schmidt (1980), Alon and Milman (1984, 1985), Jimbo and Maruoka (1985) and Buck (1986). The Jimbo–Maruoka method uses only elementary but rather complicated tools from linear algebra. The other authors apply either results from group representations or from harmonic analysis. Some of these constructions supplied better superconcentrators of densities 261.5 (Chung 1978), 218 (Jimbo and Maruoka 1985), and 122.7 (Alon et al. 1987).

More recently, Lubotzky et al. (1986, 1988) and independently Margulis (1988), applied some results of Eichler and Igusa on the Ramanujan conjectures and constructed, for every prime $p \equiv 1 \,(\mathrm{mod}\, 4)$, an infinite family of $d = (p+1)$-regular graphs G_n with $\lambda(G_i) \geqslant d - 2\sqrt{d-1}$. It is not difficult to show (see Alon 1986a or Lubotzky et al. 1988), that this is best possible. Let us describe some of these strong expanders, called Ramanujan Graphs, summarize their properties and discuss, very briefly, their connection to the Ramanujan conjectures.

Let p and q be unequal primes, both congruent to 1 modulo 4. As usual, denote by $\mathrm{PGL}(2,\mathbb{Z}_q)$ the factor group of the group of all two by two invertible matrices over $\mathrm{GF}(q)$ modulo its normal subgroup consisting of all scalar matrices. Similarly, denote by $\mathrm{PSL}(2,\mathbb{Z}_q)$ the factor group of the group of all two by two matrices over $\mathrm{GF}(q)$ with determinant 1 modulo its normal subgroup consisting of the two scalar matrices $\left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)$ and $\left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)$. The graphs we describe are $(p+1)$-regular Cayley graphs of $\mathrm{PSL}(2,\mathbb{Z}_q)$ in case p is a quadratic residue modulo q and of $\mathrm{PGL}(2,\mathbb{Z}_q)$ in case p is a quadratic nonresidue. A well-known theorem of Jacobi asserts that the number of ways of representing a positive integer n as a sum of 4 squares is

$$r_1(n) = 8\sum_{\substack{d|n\\ 4\nmid d}} d\,.$$

This easily implies that there are precisely $p+1$ vectors $a=(a_0,a_1,a_2,a_3)$, where a_0 is an odd positive integer, a_1,a_2,a_3 are even integers and $a_0^2+a_1^2+a_2^2+a_3^2=p$. Associate each such vector with the matrix γ_a in $\mathrm{PGL}(2,\mathbb{Z}_q)$ where

$$\gamma_a = \begin{bmatrix} a_0+ia_1 & a_2+ia_3 \\ -a_2+ia_3 & a_0-ia_1 \end{bmatrix},$$

and i is an integer satisfying $i^2 \equiv -1 \pmod{q}$. If p is a quadratic residue modulo q, all these matrices lie in the index two subgroup $\mathrm{PSL}(2,\mathbb{Z}_q)$ of $\mathrm{PGL}(2,\mathbb{Z}_q)$. In this case, let $G^{p,q}$ denote the Cayley graph of $\mathrm{PSL}(2,\mathbb{Z}_q)$ with respect to these $p+1$ matrices. If p is a quadratic non-residue modulo q, let $G^{p,q}$ denote the Cayley graph of $\mathrm{PGL}(2,\mathbb{Z}_q)$ with respect to the above matrices. The properties of the graphs $G^{p,q}$ are summarized in the following theorem. A detailed proof appears in Lubotzky et al. (1988).

Theorem 3.4. (i) *If p is a quadratic non-residue then $G^{p,q}$ is a bipartite $d=(p+1)$-regular graph with $n=q(q^2-1)$ nodes. Its girth is at least $4\log_p q-\log_p 4$ and its diameter is at most $2\log_p n+2\log_p 2+1$. The eigenvalues of the adjacency matrix of $G^{p,q}$, besides $(p+1)$ and $-(p+1)$, are all in absolute value at most $2\sqrt{p}$. In particular*

$$\lambda(G^{p,q}) \geqslant p+1-2\sqrt{p}=d-2\sqrt{d-1}\,.$$

(ii) *If p is a quadratic residue modulo q, then $G^{p,q}$ is a $d=(p+1)$-regular graph on $n=q(q^2-1)/2$ nodes. Its girth is at least $2\log_p q$ and its diameter is at most $2\log_p n+2\log_p 2+1$. The maximum independent set of nodes of $G^{p,q}$ is of size at most $(2\sqrt{p}/(p+1+2\sqrt{p}))n$ and its chromatic number is at least $1+(p+1)/2\sqrt{p}$. Each eigenvalue of the adjacency matrix of $G^{p,q}$, besides $p+1$, is, in absolute value, at most $2\sqrt{p}$. Hence $\lambda(G^{p,q}) \geqslant p+1-2\sqrt{p}=d-2\sqrt{d-1}$.*

Most of the properties of the graphs $G^{p,q}$ stated above are consequences of their spectral properties, i.e., the bound on the absolute values of their eigenvalues. These bounds are obtained by applying results of Eichler and Igusa concerning the Ramanujan conjectures (see Ramanujan 1916). Eichler's proof makes use of Weil's "Riemann hypothesis for curves" mentioned in the next section. These results supply good approximation for the number of ways a positive integer can be represented as a sum of four squares of a certain type. Specifically, let $r_q(n)$ denote the number of integral solutions of the quadratic equation $x_1^2+4q^2x_2^2+4q^2x_3^2+4q^2x_4^2=n$. Jacobi's theorem mentioned above, determines $r_1(n)$ precisely. For general q and for $n=p^k$, $k \geqslant 0$, there is no precise formula but the Ramanujan conjectures (which is known in this case by Eichler and Igusa's results) states that for every $\varepsilon > 0$ as k tends to infinity,

$$r_q(p^k)=C(p^k)+\mathrm{O}_\varepsilon(p^{(\frac{1}{2}+\varepsilon)k}) , \tag{3.1}$$

where $C(p^k)$, which is the main term, has an explicit known formula. In order to establish the spectral properties of the graph $G^{p,q}$, one obtains an expression for $r_q(p^k)$ in terms of the eigenvalues of $G^{p,q}$. Comparing this expression with (3.1) the desired bounds for the eigenvalues follow. The details appear in Lubotzky et al. (1988).

The Ramanujan expanders are useful in constructing efficient sorting and fault-tolerant networks. In particular, they supply superconcentrators of density 58. Although this is much better than all the previous constructions it is still worse than the best non-constructive bound, due to Bassalygo (1981) who showed, using probabilistic arguments, that there are superconcentrators of density 36.

4. Character sums and pseudo-random graphs

4.1. *Weil theorem and character sums*

Let $f(x,y)$ be a polynomial of total degree d over the finite field GF(q), with N zeros (x,y) in GF(q) × GF(q). Suppose $f(x,y)$ is absolutely irreducible, i.e., irreducible over every algebraic extension of GF(q). The famous theorem of Weil (1948), known as the Riemann hypothesis for curves over finite fields, states that

$$|N-q| \leqslant 2g\sqrt{q}+c_1(d) ,$$

where $g \leqslant \binom{d-1}{2}$ is the "genus" of the curve $f(x,y)=0$, and $c_1(d)$ depends only on d.

This highly nontrivial theorem, which was already mentioned in the previous section while briefly discussing the Ramanujan conjecture, is one of the fundamental results in modern number theory. Weil's original proof relied heavily on several ideas from algebraic geometry. Twenty years later, Stepanov found a more elementary proof, related to methods in diophantine approximation, for several special cases. His method was extended by Bombieri and Schmidt who finally obtained an elementary (but complicated) proof for the general result. A detailed presentation of several results related to Weil's theorem, using the Stepanov method, appeared in Schmidt (1976). Weil's theorem implies several sharp estimates for character sums. For our purposes here we state one, whose proof by the Stepanov method can be found in Schmidt (1976, see Theorem 2.C', p. 43).

Theorem 4.1. *Let χ be a multiplicative character of order $m > 1$ of* GF(q), *and suppose $f(x)$ has d distinct zeros in the algebraic closure of* GF(q) *and is not an mth power. Then*

$$\left|\sum_{x\in \mathrm{GF}(q)} \chi(f(x))\right| \leqslant (d-1)q^{1/2} .$$

Graham and Spencer (1971) applied this theorem to establish a pseudo-random property of a properly defined tournament. This is described below.

4.2. *Schütte's problem*

A *tournament* T_n on n nodes is an orientation of the complete graph on n nodes. For two nodes x, y of T_n, we say that x *dominates* y if the edge between x and y is directed from x to y. K. Schütte asked in 1962 whether for every $k > 0$ there is a tournament $T = T_{n(k)}$ such that for every set S of k nodes of T there is a node y which dominates all elements of S. Erdős (1963) showed, by probabilistic arguments, that for each k such a $T_{n(k)}$, with $\mathrm{O}(k^2 \cdot 2^k)$ nodes, exists. Graham and Spencer (1971) gave an explicit construction of such a tournament with $\mathrm{O}(k^2 2^{2k})$ nodes. In fact, their construction was not new; these tournaments, known as the quadratic residue or Paley tournaments, were studied before. The novelty was the application of Theorem 4.1 that showed that these tournaments have the desired properties.

4.3. *The construction*

Let q be an odd prime power congruent to 3 modulo 4. Let T_q be a tournament whose nodes are the elements of the finite field GF(q), and an edge is directed from x to y if and only if $x - y$ is a square in GF(q). Since -1 is not a square, T_q is a well-defined tournament.

Theorem 4.2. *If $q > k^2 \cdot 2^{2k-2}$ then for every set S of k nodes of T_q there is a node y which dominates all elements of S.*

Proof. Let $A = \{a_1, a_2, \ldots, a_k\}$ be a set of k distinct elements of GF(q). Let χ be the quadratic character on GF(q), i.e., for $y \in \mathrm{GF}(q)$, $\chi(y) = 1$ if y is a nonzero square in GF(q), $\chi(y) = -1$ if y is nonsquare, and $\chi(0) = 0$. We must show that there is a $y \in \mathrm{GF}(q)\backslash A$ such that $\chi(y - a_i) = 1$ for $1 \leqslant i \leqslant k$. Define

$$g(A) = \sum_{y\in \mathrm{GF}(q)\backslash A} \prod_{j=1}^{k} (1 + \chi(y - a_j)) .$$

Clearly, it is enough to show that $g(A) > 0$, since in this case there is a $y = y_0 \notin A$ such that $\prod_{j=1}^{k} (1 + \chi(y - a_j)) > 0$. To show that $g(A) > 0$, define $h(A)$ by

$$h(A) = \sum_{y\in \mathrm{GF}(q)} \prod_{j=1}^{k} (1 + \chi(y - a_j))$$

and notice that

$$g(A) = h(A) - \sum_{i=1}^{k} \prod_{j=1}^{k} \left(1 + \chi(a_i - a_j)\right) . \tag{4.1}$$

Expanding the expression for $h(A)$ we obtain

$$h(A) = \sum_{y \in \mathrm{GF}(q)} 1 + \sum_{y \in \mathrm{GF}(q)} \sum_{j=1}^{k} \chi(y - a_j) + \cdots$$
$$+ \sum_{y \in \mathrm{GF}(q)} \sum_{1 \leqslant j_1 < \cdots < j_k \leqslant k} \chi(y - a_{j_1}) \cdots \chi(y - a_{j_k}) .$$

The first two terms here are q and 0, respectively. By Theorem 4.1

$$\left| \sum_{y \in \mathrm{GF}(q)} \sum_{j_1 < \cdots < j_s} \chi(y - a_{j_1}) \cdots \chi(y - a_{j_s}) \right| \leqslant \binom{k}{s} \cdot (s-1) \cdot q^{1/2} ,$$

and hence

$$|h(A) - q| \leqslant q^{1/2} \sum_{s=2}^{k} \binom{k}{s} \cdot (s-1) = q^{1/2}\left((k-2)2^{k-1} + 1\right) .$$

Thus

$$h(A) \geqslant q - \left((k-2)2^{k-1} + 1\right)q^{1/2} .$$

Using (4.1) one can easily check that $h(A) - g(A) \leqslant 2^{k-1}$ and thus, if $q > k^2 2^{2k-2}$, $g(A) > 0$. This completes the proof. □

4.4. The pseudo random properties of T_q

An easy variation of the last proof shows that the tournament T_q constructed above has the following property: For every two disjoint sets of nodes A, B of T_q, with $|A| + |B| = k$, the number of nodes y of T_q that dominate all members of A and are dominated by all members of B is

$$\frac{q}{2^k} + \mathrm{O}(k \cdot q^{1/2} + k \cdot 2^k) .$$

Thus, for say, $k < \frac{1}{4} \log q$, this number is very close to $(q-k)/2^k$ which is the expected number of such nodes in a random tournament on q nodes. This easily implies that the number of labeled subtournaments of T_q isomorphic to any given labeled tournament on k nodes is very close to $n/2^{\binom{k}{2}}$. Thus T_q resembles a random tournament on q nodes.

As observed by Bollobás and Thomason (1981), a similar construction supplies undirected pseudo-random graphs. These are called Paley Graphs. Suppose q is an odd prime power, congruent to 1 modulo 4, and let G_q be the graph whose nodes are the elements of GF(q), where x and y are adjacent if $x - y$ is a square in GF(q). As before, one can show, using Theorem 4.1, that for every two disjoint sets A, B of nodes, with $|A| + |B| = k$, $k < \frac{1}{4}\log q$, the number of nodes y adjacent to all elements of A and nonadjacent to every element of B is very close to $q/2^k$. This implies, of course, that G_q contains all graphs on k vertices as induced subgraphs.

It seems more difficult to construct explicitly large graphs that do not contain some specified small induced subgraphs. In fact, the best known open problem concerning explicit constructions is a problem of this type. This is the problem of obtaining constructive lower bounds for the usual diagonal Ramsey numbers. Specifically, we want to describe explicitly, for every k, a graph with c^k nodes that contains neither a clique of size k, nor a stable set of size k, where $c > 1$ is a constant independent of k. The best known result in this direction is that of Frankl and Wilson (1981) who constructed such a graph with $\exp(\Omega(\log^2 k/\log\log k))$ nodes. It may be true that for primes q the Paley graphs G_q are better examples, but at present a proof of this, which would have several new number-theoretic consequences, seems hopeless.

5. Real varieties and sign patterns of polynomials

5.1. The number of connected components

In this section we describe several combinatorial applications of the known estimates for the number of connected components of real varieties or semivarieties. Such estimates were obtained by several authors, and can be found, among other places, in Oleĭnik and Petrovskiĭ (1949), Milnor (1964), Thom (1965) and Warren (1968). For our purposes all these existing bounds suffice. To be specific, we state two of them.

Theorem 5.1 (Milnor 1964). *Let V be a real variety in $\mathbb{R}^\ell$, defined by the solution set of the real polynomial equations*

$$f_i(x_1, \ldots, x_\ell) = 0\ , \quad i = 1, \ldots, m\ ,$$

and suppose the degree of each polynomial f_i is at most k. Then the number $c(V)$ of connected components of V is at most $k \cdot (2k-1)^{\ell-1}$.

Theorem 5.2 (Warren 1968). *Let $P_1, \ldots, P_m$ be real polynomials in ℓ variables, each of degree k or less. Let V be the set $\{\underline{x} \in \mathbb{R}^\ell \colon P_i(\underline{x}) \neq 0$ for all $1 \leqslant i \leqslant m\}$. Then the number $c(V)$ of connected components of V does not exceed $2(2k)^\ell \sum_{i=0}^{\ell} 2^i \binom{m}{i}$. In particular, if $m \geqslant \ell \geqslant 2$ then*

$$c(V) \leqslant (4ekm/\ell)^\ell\ .$$

We note that Theorem 5.1 can be applied to deduce upper bounds for the number of connected components of the solution set of a system of algebraic inequalities, by expressing such a set as a projection of a variety in a higher dimension.

5.2. Lower bounds for algebraic decision trees

In an elegant paper, Steele and Yao (1982) applied Milnor's result stated above to obtain lower bounds for the height of algebraic decision trees. Their method was modified and extended by Ben-Or (1983). We outline this method below. Related interesting results appear in Björner et al. (1992).

For $W \subseteq \mathbb{R}^\ell$, the *membership problem for* W is the following:

Given $\underline{x} = (x_1, \ldots, x_\ell) \in \mathbb{R}^\ell$, determine if $\underline{x} \in W$. Thus, for example, the *ℓ-element distinctness problem*, which is the problem of deciding whether ℓ given real numbers are all distinct, is just the membership problem for

$$W = \left\{ (x_1, \ldots, x_\ell) \in \mathbb{R}^\ell : \prod_{1 \leqslant i < j \leqslant \ell} (x_i - x_j) \neq 0 \right\} .$$

We are interested in algorithms for solving the membership problem for W that allow arithmetic operations and tests. More formally, an *algebraic decision* tree is a binary tree T with a rule that assigns:

(a) To any node v with one son, an operational instruction of the form:

$$f_v = f_{v_1} \circ f_{v_2} \quad \text{or} \quad f_v = c \circ f_{v_1} ,$$

where v_i is an ancestor of v in T, or $f_{v_i} \in \{x_1, \ldots, x_\ell\}$, $\circ \in \{+, -, \times, /\}$ and $c \in \mathbb{R}$.

(b) To any vertex v with two sons, a test instruction of the form $f_{v_1} > 0$ or $f_{v_1} \geqslant 0$ or $f_{v_1} = 0$, where v_1 is an ancestor of v or $f_{v_1} \in \{x_1, \ldots, x_\ell\}$.

(c) To any leaf an output Yes or No.

For any input $x \in \mathbb{R}^\ell$, the algorithm traverses a path $P(x)$ in T from the root, where at each node, the corresponding arithmetic operation is performed or a branching is made according to the test. When a leaf is reached, the anwser Yes or No to the problem is returned. We note that one can allow more algebraic operations (like square roots, etc.), but the treatment is similar.

Theorem 5.3 (Ben-Or 1983). *Suppose $W \subseteq \mathbb{R}^\ell$, and let T be an algebraic decision tree that solves the membership problem for W (i.e., for each $\underline{x} \in \mathbb{R}^\ell, P(\underline{x})$ ends in a "Yes" leaf iff $\underline{x} \in W$). If h is the height of T then*

$$2^h \cdot 3^{\ell+h} \geqslant N ,$$

where N is the number of connected components of W.

Proof. The main tool in the proof of this theorem is Theorem 5.1 stated above. One first observes that every "Yes" leaf corresponds to a subset of $\mathbb{R}^\ell$ that is a projection of a variety defined by a system of at most h quadratic equations and inequalities in $\ell + h$ variables. Using Theorem 5.1 one can show that such a subset can have at most $3^{\ell+h}$ connected components. Each such component must be contained in some connected component of W, and since the number of leaves of T is at most 2^h, and all components of W must be covered by the "Yes" leaves we have $2^h \cdot 3^{\ell+h} \geqslant N$. □

As an example for applying Theorem 5.3, notice that

$$W = \left\{ (x_1,\dots,x_\ell) \in \mathbb{R}^\ell \colon \prod_{1 \leqslant i<j \leqslant \ell} (x_i - x_j) \neq 0 \right\}$$

has precisely $\ell!$ connected components, corresponding to the $\ell!$ possible order-types of $x_1,\dots,x_\ell$. Thus, any algebraic decision tree that solves the ℓ-elements distinctness problem has height $\Omega(\ell \log \ell)$. This is clearly best possible, as the ℓ-element distinctness problem can be solved by sorting the ℓ elements and then comparing all pairs of adjacent elements in the sorted order.

5.3. Sign patterns of real polynomials

For further applications of Theorems 5.1 and 5.2, it will be convenient to derive a more combinatorial corollary, dealing with sign patterns of real polynomials.

Let $P_j = P_j(x_1,\dots,x_\ell)$, $j = 1,\dots,m$ be m real polynomials. For a point $\underline{c} \in \mathbb{R}^\ell$, the *sign-pattern* of the P_j's at $\underline{c}$ is the m-tuple $(\varepsilon_1,\dots,\varepsilon_m) \in (-1,0,1)^m$, where $\varepsilon_j = \operatorname{sign} P_j(\underline{c})$. Let $s(P_1,P_2,\dots,P_m)$ denote the total number of sign-patterns of the polynomials $P_1,P_2,\dots,P_m$, as $\underline{c}$ ranges over all points of $\mathbb{R}^\ell$. Similarly, let $\bar{s}(P_1,P_2,\dots,P_m)$ denote the total number of sign-patterns consisting of vectors with $\{\pm 1\}$ coordinates. Clearly, $s(P_1,P_2,\dots,P_m) \leqslant 3^m$ and $\bar{s}(P_1,P_2,\dots,P_m) \leqslant 2^m$. However, one can apply Theorem 5.1 or Theorem 5.2 to bound these numbers by a function of ℓ and the degrees of the polynomials $P_1,P_2,\dots,P_m$. Indeed, suppose the degree of each P_i does not exceed k. Put $V = \{\underline{x} \in \mathbb{R}^\ell \colon P_i(\underline{x}) \neq 0 \quad \text{for all} \quad 1 \leqslant i \leqslant m\}$. Clearly $\bar{s}(P_1,\dots,P_m)$ is bounded above by the number $c(V)$ of connected components of V. This, together with Theorem 5.2, gives the following result (for $\ell \geqslant 2$; for $\ell = 1$ it is trivial).

Proposition 5.4 (Warren 1968). *Let $P_1,\dots,P_m$ be m real polynomials in ℓ real variables, and suppose the degree of each P_i does not exceed k. If $m \geqslant \ell$ then $\bar{s}(P_1,P_2,\dots,P_m) \leqslant (4ekm/\ell)^\ell$.*

It is not too difficult to obtain a similar bound for the total number $s(P_1,P_2,\dots,P_m)$ of sign-patterns. Indeed, let $C \subseteq \mathbb{R}^\ell$ be a set of cardinality $|C| = s(P_1,P_2,\dots,P_m)$ representing all sign-patterns of the polynomials $P_1,P_2,\dots,P_m$. Define $\varepsilon > 0$ by

$$\varepsilon = \tfrac{1}{2} \min\{|P_j(\underline{c})| \colon \underline{c} \in C, 1 \leqslant j \leqslant m \text{ and } P_j(\underline{c}) \neq 0\} .$$

Now put $V = \{\underline{x} \in \mathbb{R}^\ell \colon P_i(\underline{x}) - \varepsilon \neq 0 \quad \text{and} \quad P_i(\underline{x}) + \varepsilon \neq 0 \quad \text{for all} \quad 1 \leqslant i \leqslant m\}$. Clearly $C \subseteq V$ and one can easily check that each two distinct points $\underline{c}, \underline{c}' \in C$ lie in distinct connected components of V. Hence $s(P_1,\dots,P_m) = |C|$ does not exceed the number of connected components of V. In view of Theorem 5.2, we conclude the following.

Proposition 5.5. *Let $P_1,\dots,P_m$ be m real polynomials in ℓ real variables, and suppose the degree of each P_j does not exceed k. If $2m \geqslant \ell$ then $s(P_1,\dots,P_m) \leqslant (8ekm/\ell)^\ell$.*

A similar estimate can be obtained from Theorem 5.1.

5.4. *The number of polytopes and configurations*

If $(P_0, P_1, \ldots, P_d)$ is a sequence of $d+1$ points in R^d, with $P_i = (x_{i1}, \ldots, x_{id})$ for each i, we say they have *positive orientation*, written $P_0 \ldots P_d > 0$, if $\det(x_{ij})_{0<i,j<d} > 0$ where $x_{i0} = 1$ for each i. The conditions $P_0 \ldots P_d < 0$ and $P_0 \ldots P_d = 0$ are defined similarly. The *order type* of a configuration C of n labeled points $P_1, P_2, \ldots, P_n$ in $\mathbb{R}^d$ is a function w from the set of all $(d+1)$-subsets of $\{1,2,\ldots,n\}$ to $\{0, \pm 1\}$, where for $S = \{i_0, i_1, \ldots, i_d\}$ with $1 \leqslant i_0 < i_1 < \cdots < i_d \leqslant n, w(S) = +1$ if $P_{i_0} \ldots P_{i_d} > 0$, $w(S) = -1$ if $P_{i_0} \ldots P_{i_d} < 0$, and $w(S) = 0$ if $P_{i_0} \ldots P_{i_d} = 0$. The configuration is *simple* if $w(S) \neq 0$ for every such S. Notice that $w(S)$ is just sign $\det(x_{i_k j})$, $0 \leqslant k, j \leqslant d$, where $P_{i_k} = (x_{i_k 1}, \ldots, x_{i_k d})$ and $x_{i_k 0} = 1$ for $0 \leqslant k \leqslant d$. The order type of a configuration C of points is sometimes known as the oriented matroid structure determined by C. Let $t(n,d)$ denote the number of distinct order types of configurations of n labeled points in $\mathbb{R}^d$, and let $t_s(n,d)$ denote the number of order types of simple configurations of n labeled points in $\mathbb{R}^d$. Goodman and Pollack (1986) applied Milnor's theorem 5.1 to show that $t_s(n,d) \leqslant n^{d(d+1)n}$. As it is not too difficult to show that for every fixed $d \geqslant 2$, $t_s(n,d) \geqslant n^{(1+o(1))d^2 n}$, as n tends to infinity, this upper bound is not far from the truth. In Alon (1986b) it is shown that $n^{(1+o(1))d^2 n}$ is the correct order of magnitude of both $t_s(n,d)$ and $t(n,d)$. This is, in fact, an immediate consequence of Proposition 5.5.

Proposition 5.6. *For every fixed $d \geqslant 2$, as n tends to infinity,*

$$t_s(n,d) \leqslant t(n,d) \leqslant n^{(1+o(1))d^2 n} .$$

Proof. Obviously $t(n,d)$ is just the number of sign patterns of $\binom{n}{d+1}$ polynomials of degree d in the dn real variables $(x_{i1}, \ldots, x_{id})$, which are the coordinates of the ith point $(1 \leqslant i \leqslant n)$. The polynomials are just all the determinants $\det(x_{i_k j}), 0 \leqslant k, j \leqslant d$, where $x_{i_k 0} = 1$ for all k and $1 \leqslant i_0 < i_1 < \cdots < i_d \leqslant n$. The result now follows from Proposition 5.5. □

The same computation shows that for every n and d

$$t_s(n,d) \leqslant t(n,d) \leqslant 2^{n^3 + O(n^2)} .$$

Next we consider the number of combinatorial types of convex polytopes.

Let $c(n,d)$ denote the number of (combinatorial types of) d-polytopes on n labeled vertices and let $c_s(n,d)$ denote the number of simplicial d-polytopes on n labeled vertices. The problem of determining or estimating these two functions (especially for 3-polytopes) was the subject of much effort and frustration of nineteenth-century geometers. Although it follows from Tarski's theorem on the decidability of first order sentences in the real field that the problem of computing $c(n,d)$ is solvable, it seems extremely difficult to actually determine this number even for relatively small n and d. Both Cayley and Kirkman failed to determine $c(n,3)$ or $c_s(n,3)$ despite a lot of effort. Detailed historical surveys of these attempts were given by Brückner and Steinitz (cf. Grünbaum 1967, pp. 288–290), and the asymptotic behaviour of $c(n,d)$ and $c_s(n,d)$ is known only for $d \leqslant 3$ or $n \leqslant d+3$. It is thus pleasing to note, following Goodman and Pollack (1986) and, later, Alon (1986b) that Proposition 5.5 supplies immediately upper bounds for $c_s(n,d)$ and for $c(n,d)$. This follows from the fact that the order type of a configuration that spans $\mathbb{R}^d$ determines which sets of its points lie on supporting hyperplanes of its convex hull. Hence, the order type of a configuration of a set of n points in $\mathbb{R}^d$ which is the set of vertices of a convex polytope P determines its facets and its complete combinatorial type. Thus Proposition 5.6 and the paragraph following it imply the following result.

Proposition 5.7. *For every fixed d,*

$$c_s(n,d) \leqslant c(n,d) \leqslant n^{(1+o(1))d^2 n} .$$

Furthermore, the total number of polytopes of any dimension on n points is at most $2^{n^3+O(n^2)}$.

Although this proposition is an immediate corollary of the known bounds for sign-patterns of polynomials, it improves considerably the previously best known bound which was $n^{O(n^{d/2})}$. We note also that one can show that for every $n \geqslant 2d$,

$$c_s(n,d) \geqslant \left(\frac{n-d}{d}\right)^{nd/4} .$$

5.5. *Ranks of sign matrices*

The *sign-pattern* of an m by n matrix A with nonzero entries $(a_{ij})_{1\leqslant i\leqslant m, 1\leqslant j\leqslant n}$ is an m by n matrix $Z(A) = (x_{ij})$ of $1, -1$ entries where $z_{ij} = \text{sign } a_{ij}$. For an m by n matrix Z of $1, -1$ entries, let $r(Z)$ be the minimum possible rank of a matrix A such that $Z(A) = Z$. Define $r(n,m) = \max\{r(Z)$: Z is an m by n matrix over $\{1,-1\}\}$. The problem of determining or estimating $r(n,m)$, and in particular $r(n,n)$, was raised by Paturi and Simon (1984). They observed that $r(n,n) \geqslant \lfloor \log_2 n \rfloor$ and raised the question if one can prove a superlogarithmic lower bound for $r(n,n)$. As shown in their paper, this would supply lower bounds for the maximal possible *unbounded-error probabilistic communication complexity* of a Boolean function of $2p$ bits. This question is answered in Alon et al. (1985) where it is shown, in particular, that

$$\frac{n}{16} \leqslant r(n,n) \leqslant \frac{n}{2} + 3\sqrt{n} ,$$

and that if $m/n^2 \to \infty$ and $(\log_2 m)/n \to 0$ then

$$r(n,m) = \left(\tfrac{1}{2} + o(1)\right) n .$$

(The bounds here are slightly better than those that actually appear in the above paper.)

The upper bounds are proved by combining some simple geometric, combinatorial and probabilistic arguments. The lower bounds can be deduced, by a simple counting argument, from Propositions 5.4 and 5.6.

As shown in Alon et al. (1985), these results imply that the (bounded or unbounded)-error probabilistic communication complexity of almost every Boolean function of $2p$ variables is between $p-4$ and p.

5.6. *Degrees of freedom versus dimension of containment orders*

The *dimension* of a partially ordered set P is the minimum number of linear extensions whose intersection is P. Alternatively, it is the smallest k so that the elements of P can be mapped to points in $\mathbb{R}^k$ so that $x \leqslant y$ iff each coordinate of x is less than or equal to the corresponding coordinate of y.

Let $\mathcal{S}$ be a family of sets. We say that a partially ordered set P has an $\mathcal{S}$-*containment representation* provided there is a map $f : P \to \mathcal{S}$ such that $x < y$ iff $f(x) \subset f(y)$. In this case we say that P is an $\mathcal{S}$-order.

For example, *circle orders* are the containment orders of disks in the plane. Similarly, angle orders are the containment orders of angles in the plane, where an angle includes its interior.

Note that circles admit three "degrees of freedom": two center coordinates and a radius. An angle admits four degrees of freedom: the two coordinates of its vertex and the slopes of its rays. Further, it is known that not all 4-dimensional posets are circle orders nor are all 5-dimensional posets angle orders. These are confirming instances of the following intuitive notion: *If the sets in $\mathcal{S}$ admit k degrees of freedom, then not all $(k+1)$-dimensional posets are $\mathcal{S}$-orders.*

Let us briefly show, following Alon and Scheinerman (1988), how the estimates for the number of sign patterns of real polynomials, supply a precise version of this intuitive principle. We say that the sets in $\mathcal{S}$ have k *degrees of freedom* provided:

(1) Each set in $\mathcal{S}$ can be uniquely identified by a k-tuple of real numbers, i.e., there is an injection $f : \mathcal{S} \to \mathbb{R}^k$.

(2) There exists a finite list of polynomials $p_1, p_2, \ldots, p_t$ in $2k$ variables with the following property: If $S, T \in \mathcal{S}$ map to $(x_1, \ldots, x_k), (y_1, \ldots, y_k) \in \mathbb{R}^k$, respectively, then the containment $S \subset T$ can be determined based on the signs of the values $p_j(x_1, \ldots, x_k, y_1, \ldots, y_k)$ for $1 \leqslant j \leqslant t$.

For example, let us consider disks in the plane. Suppose we have two disks C_1 and C_2 with centers and radii given by x_i, y_i, r_i $(i = 1,2)$. One checks that we have $C_1 \subset C_2$ iff both the following hold:

$$(x_1 - x_2)^2 + (y_1 - y_2)^2 - (r_1 - r_2)^2 \leqslant 0\,,$$
$$r_1 - r_2 \leqslant 0\,.$$

Thus the family of circles in the plane admits three degrees of freedom. Similarly, the containment of one angle in another can be expressed in terms of a finite list of polynomial inequalities.

Theorem 5.8. *Let $\mathcal{S}$ be a family of sets admitting k degrees of freedom. Then the number of $\mathcal{S}$-orders on n labeled points is at most*

$$2^{(1+o(1))kn\log n} ,$$

as n tends to infinity.

Proof. Let $\mathcal{S}_n$ denote the family of $\mathcal{S}$-orders on $\{1,\dots,n\}$. For each n-tuple of sets in $\mathcal{S}$, $(S_1,\dots,S_n)$ we have a (potentially) different poset depending on the sign pattern of $r = 2\binom{n}{2}t$ polynomials in $\ell = nk$ variables which have some maximum degree d (which is independent of n). Hence by Proposition 5.5,

$$|\mathcal{S}_n| \leqslant \left[\frac{16ed\binom{n}{2}t}{nk}\right]^{nk} = [O(1)n]^{nk} = 2^{(1+o(1))kn\log n} . \qquad \square$$

Denote by $P(n,k)$ the number of posets of dimension at most k on n labeled points $\{1,2,\dots,n\}$. By a simple construction, one can show that for every fixed k

$$\lim_{n\to\infty} \log P(n,k)/(kn\log n) = 1 .$$

This and the previous proposition imply the following.

Corollary 5.9. *Let $\mathcal{S}$ be a family of sets admitting k degrees of freedom. Then there exists a $(k+1)$-dimensional poset which is not an $\mathcal{S}$-containment order.*

6. The Chevalley–Warning theorem, abelian groups and regular graphs

The classical theorem of Chevalley and Warning, that deals with the number of solutions of a system of polynomials with many variables over a finite field, is the following.

Theorem 6.1 (see, e.g., Borevich and Shafarevich 1966 or Schmidt 1976). *For $j = 1,2,\dots,n$ let $P_j(x_1,\dots,x_m)$ be a polynomial of degree r_j over a finite field F of characteristic p. If $\sum_{j=1}^{n} r_j < m$ then the number N of common zeros of $P_1,\dots,P_n$ (in F^n) satisfies*

$$N \equiv 0 \pmod p .$$

In particular, if there is one common zero, then there is another one.

Proof. The proof is extremely simple; clearly, if F has q elements then

$$N \equiv \sum_{x_1,\dots,x_m\in F} \prod_{j=1}^{n} \left(1 - P_j(x_1,\dots,x_m)^{q-1}\right) \pmod p . \tag{6.1}$$

By expanding the right-hand side we get a linear combination of monomials of the form

$$\prod_{i=1}^{m} x_i^{k_i} \quad \text{with} \quad \sum_{i=1}^{m} k_i \leqslant (q-1)\sum_{j=1}^{n} r_j < (q-1)m .$$

Hence, in each such monomial there is an i with $k_i < q - 1$. But then in $F = \mathrm{GF}(q)$, $\sum_{x_i \in F} x_i^{k_i} = 0$, implying that the contribution of each monomial to the sum (6.1) is $0 \pmod{p}$, completing the proof. □

In this section we discuss some applications of this theorem to combinatorial problems in abelian groups, extremal graph theory and the theory of finite affine spaces.

6.1. Combinatorial problems in abelian groups

For a finite abelian group G, define $s = s(G)$ to be the smallest positive integer such that, for any sequence $g_1, g_2, \ldots, g_s$ of (not necessarily distinct) elements of G, there is an $\phi \neq I \subset \{1, \ldots, s\}$ such that $\sum\{g_i : i \in I\} = 0$. The problem of determining $s(G)$ was proposed by H. Davenport in 1966, and is related to the study of the maximal number of prime ideals in the decomposition of an irreducible integer in an algebraic number field whose class group is G. Olson (1969a) determined $s(G)$ for every p-group $G = \mathbb{Z}_p e_1 \oplus \cdots \oplus \mathbb{Z}_p e_r$. Clearly

$$s(G) \geqslant 1 + \sum_{i=1}^{r} (p^{e_i} - 1) ,$$

for let $x_1, \ldots, x_r$ be a basis for G, where x_i has order p^{e_i}, and consider the sequence of length $\sum_{i=1}^{r} (p^{e_i} - 1)$ in which each x_i occurs $p^{e_i} - 1$ times. No subsequence here has sum 0. Olson gave a charming proof of the following.

Theorem 6.2. $s(\mathbb{Z}_p e_1 \oplus \cdots \oplus \mathbb{Z}_p e_r) = 1 + \sum_{i=1}^{r} (p^{e_i} - 1)$.

For the case $e_1 = \cdots = e_r = 1$ this can be easily deduced from the Chevalley–Warning theorem as follows. Let $g_1, g_2, \ldots, g_s$ be a sequence of elements of $G = (\mathbb{Z}_p)^r$, where $s > r(p-1)$ and put $g_i = (g_{i1}, g_{i2}, \ldots, g_{ir})$. Consider the following system of r polynomials in s variables over $\mathrm{GF}(p)$;

$$\sum_{i=1}^{s} g_{ij} x_i^{p-1} = 0, \quad j = 1, \ldots, r .$$

Since $s > r \cdot (p-1)$ and $x_1 = \cdots = x_s = 0$ is a trivial solution, there is a nontrivial solution $(z_1, \ldots, z_s)$. Put $I = \{i : z_i \neq 0\}$ and observe that $\sum\{g_i : i \in I\} = 0$ to complete the proof (for the case $e_1 = \cdots = e_r = 1$).

The general case can be proved by generalizing the proof of the Chevalley–Warning theorem.

Olson's original proof is different and is based on the fact that the ideal of nilpotent elements in the group-ring of a p-group over $\mathbb{Z}_p$ is nilpotent. Since this proof is short and elegant, we present it in full.

Proof of Theorem 6.2. Let G be the finite abelian p-group with invariants $p^{e_1}, p^{e_2}, \ldots, p^{e_r}$, and let us use multiplicative notation for G. Let R be the group ring of G over $\mathbb{Z}_p$. Suppose $k \geqslant 1 + \sum_{i=1}^{r}(p^{e_i} - 1)$ and let $g_1, g_2, \ldots, g_k$ be a sequence of k members of G. We claim that in R

$$(1 - g_1) \cdot (1 - g_2) \cdots (1 - g_k) = 0 \ . \tag{6.2}$$

Indeed, let $x_1, x_2, \ldots, x_r$ be the standard basis for G, where the order of x_i is p^{e_i}. Since each g_j can be written as a product of the elements x_i, a repeated application of the identity $1 - uv = (1 - u) + u(1 - v)$ enables us to express each expression of the form $1 - g_j$ as a linear combination (with coefficients in R) of the elements $1 - x_i$. Substituting into (6.2) and applying commutativity we conclude that the left-hand side is a linear combination of elements of the form

$$\prod_{i=1}^{r}(1 - x_i)^{k_i} \ , \quad \text{where} \sum_{i=1}^{r} k_i = k > \sum_{i=1}^{r}(p^{e_i} - 1) \ .$$

Hence, there is an i with $k_i \geqslant p^{e_i}$ and since in R, $(1 - x_i)^{p^{e_i}} = 1 - x_i^{p^{e_i}} = 0$ this implies that (6.2) holds as claimed.

By interpreting (6.2) combinatorially we conclude that there is some nontrivial subsequence of $g_1, \ldots, g_k$ that has product 1, since otherwise, the coefficient of 1 in the above product will be nonzero. Hence $s(G) = 1 + \sum_{i=1}^{r}(p^{e_i} - 1)$, as needed. □

We note that if $G = C_1 \oplus \cdots \oplus C_r$ is the direct sum of cyclic groups C_i of orders $|C_i| = c_i$, where $c_i | c_{i+1}$, then $s(G) \geqslant 1 + \sum_{i=1}^{r}(c_i - 1)$, and this inequality can be strict. Several interesting generalizations of Olson's results (including an upper estimate for $s\big((\mathbb{Z}_n)^m\big)$, appear in Baker and Schmidt (1980) and in Van Emde Boas and Kruyswijk (1969). It is, however, not known if the equality $s\big(\mathbb{Z}_n)^m\big) = m(n - 1) + 1$ holds for all m and n.

Erdős et al. (1961), showed that for any sequence $g_1, g_2, \ldots, g_{2n-1}$ of elements of a finite abelian group of order n, there exists a set $I \subset \{1, \ldots, 2n - 1\}$ of n indices such that $\sum\{g_i \colon i \in I\} = 0$. The first (and main) step of their proof is to prove the above when $G = \mathbb{Z}_p$ is the cyclic group of order p, where p is a prime. Although the proof in this case is an easy consequence of a special case of the Cauchy–Davenport lemma (see chapter 20) it is interesting to note that this fact can also be derived from the Chevalley–Warning theorem as follows. Consider the following system of two polynomials in $2p - 1$ variables over GF(p):

$$\sum_{i=1}^{2p-1} g_i x_i^{p-1} = 0 \ ,$$

$$\sum_{i=1}^{2p-1} x_i^{p-1} = 0 \ .$$

Since $2(p - 1) < 2p - 1$ and $x_1 = x_2 = \cdots = x_{2p-1} = 0$ is a solution, Theorem 6.1 implies the existence of a nontrivial solution $(z_1, \ldots, z_{2p-1})$. Since in GF(p), $y^{p-1} = 1$

if $y \neq 0$ and $0^{p-1} = 0$, $I = \{i\colon z_i \neq 0\}$ satisfies $\sum\{g_i\colon i \in I\} = 0$ and $|I| = p$, completing the proof. Notice, also, that the above result also follows from Theorem 6.2 by considering the $2p-1$ elements $(g_1, 1), (g_2, 1), \ldots, (g_{2p-1}, 1)$ in $\mathbb{Z}_p \oplus \mathbb{Z}_p$.

The following generalization of the Erdős–Ginzburg–Ziv theorem was proved by Olson (1969b).

Theorem 6.3. *Let $H = G \oplus K$ be the direct sum of the abelian groups G and K of orders $|G| = n$ and $|K| = k$, where $k|n$. If $h_1, h_2, \ldots, h_{n+k-1}$ is a sequence of $n+k-1$ elements of H, then there is a set $\phi \neq I \subset \{1, 2, \ldots, n+k-1\}$ of indices such that $\sum\{h_i\colon i \in I\} = 0$.*

This theorem can also be deduced from Theorem 6.1, together with some of the ideas of Olson (1969b). It implies the previous statement by taking K to be the cyclic group of order n and by defining $h_i = g_i \oplus 1 \in G \oplus K$ for $1 \leqslant i \leqslant 2n-1$.

6.2. *Regular subgraphs of graphs*

As shown by Alon et al. (1984), one can apply the Chevalley–Warning theorem or Olson's Theorem 6.2 to prove that certain graphs contain regular subgraphs. A graph H is *q-divisible* if q divides the degree of every node of H. Let $f(n,q)$ be the maximum number of edges of a loopless graph G on n nodes that contains no nonempty q-divisible subgraph. Suppose q is a prime power, and let $G = (V, E)$ be a loopless graph with $|V| = n$ nodes and $|E| = m > n \cdot (q-1)$ edges. Let $a_j^{(i)}$ be the (j,i)th entry of the (node–edge)-incidence matrix of G. The vectors $a^{(i)} = (a_1^{(i)}, \ldots, a_n^{(i)})$, $1 \leqslant i \leqslant m$ are elements of $(\mathbb{Z}_q)^n$, so by Olson's Theorem 6.2 there exists an $\phi \neq I \subset \{1, \ldots, m\}$ such that $\sum\{a_j^{(i)}\colon i \in I\} \equiv 0 \pmod{q}$ for $1 \leqslant j \leqslant n$. The subgraph H consisting of all edges whose indices lie in I is clearly q divisible. Hence $f(n,q) \leqslant n \cdot (q-1)$. This estimate can be slightly improved for powers of 2, and a matching lower bound can be given for all $n \geqslant 3$. Therefore, the following holds.

Theorem 6.4. *For every odd prime power q and every $n \geqslant 3$, $f(n,q) = (q-1)n$. For every power of two q and every $n \geqslant 3$, $f(n,q) = (q-1) \cdot n - \frac{1}{2}q$.*

Similarly, using the results of Van Emde Boas and Kruyswijk (1969), one can show that $f(n,k) \leqslant c(k) \cdot n$ for all k and n. The truth, however, might be that $f(n,k) \leqslant (k-1) \cdot n$ for every $n \geqslant 3$ and every k.

By Theorem 6.4, if q is a prime power and $G = (V, E)$ is a graph with maximum degree at most $2q-1$ and average degree greater than $2q-2$, then G contains a q-regular subgraph. Indeed, $|E| > (q-1) \cdot |V|$ and hence G contains a q-divisible subgraph which must be q-regular since its maximum degree is smaller than $2q$.

In particular, every loopless 4-regular graph plus an edge contains a 3-regular subgraph. This is closely related to the well-known Berge–Sauer conjecture, which asserts that every 4-regular simple graph (i.e., a graph with no loops and no parallel edges) contains a 3-regular subgraph. This conjecture has been proved by Taškinov (1982). It is, however, false for graphs with parallel edges, and hence the assumption "plus an edge" in the previous statement cannot be omitted.

Another consequence of Theorem 6.4, together with several known results in graph theory is that for every k and r that satisfy $k \geqslant 4r$, every loopless k-regular graph contains an r-regular subgraph. For several sharper results see Alon et al. (1984).

Erdős and Sauer (see, e.g., Bollobás 1978, p. 399) asked for an estimation of the maximal number of edges of a simple graph on n nodes that contains no 3-regular subgraph. They conjectured that this number is $o(n^{1+\varepsilon})$ for any $\varepsilon > 0$. This conjecture has been proved by Pyber (1985) by applying Theorem 6.4. Pyber showed that any simple graph with n nodes and at least $200 n \log n$ edges contains a subgraph with maximal degree 5 and average degree greater than 4. This subgraph contains, by the paragraph following Theorem 6.4, a 3-regular subgraph. A similar reasoning shows that there exists a constant $c > 0$ such that for every $r \geqslant 3$, every simple graph G with n nodes and at least $c \cdot r^2 \cdot n \log n$ edges contains an r-regular subgraph. On the other hand, Pyber, Rödl and Szemerédi showed, using probabilistic arguments, that there are simple graphs with n nodes and $\Omega(n \log\log n)$ edges, that contain no 3-regular subgraphs. Thus the above result is not far from being best possible.

6.3. *The blocking number of an affine space*

For a prime power q and $k > 0$, let $\mathrm{AG}(k,q)$ denote the k-dimensional affine space over $\mathrm{GF}(q)$. It is not too difficult to observe that there is always a subset of cardinality $k \cdot (q-1)+1$ that intersects all hyperplanes. Indeed, the union of any k independent lines through a point intersects all hyperplanes and has this cardinality.

Theorem 6.5. *The minimum cardinality of a subset of* $\mathrm{AG}(k,q)$ *that intersects all hyperplanes is* $k \cdot (q-1)+1$.

This theorem was proved, independently, by Jamison (1977) who gave a rather lengthy proof for a more general result, and by Brouwer and Schrijver (1978) who obtained an elegant and short proof. If $q = p$ is a prime, their proof can be shortened even further by using the Chevalley–Warning theorem as follows. Suppose $A \subset \mathrm{AG}(k,p)$ intersects all hyperplanes. We may assume that $0 = (0,\ldots,0) \in A$, and define $B = A \setminus \{0\}$. Then B intersects all hyperplanes not through 0, i.e., for every $0 \neq (w_1, w_2, \ldots, w_k) \in (\mathrm{GF}(p))^k$ there exists a $b = (b_1, \ldots, b_k) \in B$ such that $w_1 b_1 + \cdots + w_k b_k = 1$. Define $F(x_1, \ldots, x_k) = \prod_{b \in B}(1 - b_1 x_1 - \cdots - b_k x_k)$. Clearly $F(w_1, \ldots, w_k) = 0$ for all $(w_1, \ldots, w_k) \neq 0$ and $F(0, \ldots, 0) = 1$. Consider the following polynomial equation in the $k \cdot (p-1)$ variables $x_1^{(i)}, \ldots, x_k^{(i)}$, $1 \leqslant i \leqslant p-1$:

$$\sum_{i=1}^{p-1} F(x_1^{(i)}, \ldots, x_k^{(i)}) = p - 1 \; .$$

Obviously, the only zero of this equation is the trivial solution $x_j^{(i)} = 0$ for $1 \leqslant j \leqslant k$, $1 \leqslant i \leqslant p-1$. By the Chevalley–Warning theorem, this implies that the degree of the above polynomial, which is $|B|$, is at least as big as the number of variables, which is $k \cdot (p-1)$. Hence $|A| \geqslant k \cdot (p-1)+1$, as needed.

It is worth noting that neither the proof of Jamison nor the one of Brouwer and Schrijver imply any estimate for the analogous problem for non-Desarguesian planes.

7. More polynomials

In the last two sections real polynomials and polynomials over a finite field were used to derive some combinatorial results. In this section, we describe some further combinatorial problems, where polynomials and ideals of polynomials are applied for deriving certain characterization results with combinatorial consequences.

7.1. Generators of ideals, graph polynomials and vectors balancing

For a graph $G=(V,E)$ on the n nodes $\{1,2,\dots,n\}$, define the associated *graph polynomial* $f_G = f_G(x_1,\dots,x_n)$ by

$$f_G = \prod \{(x_i - x_j) \colon ij \in E\} .$$

The *independence number* $c(G)$ (= the maximum size of a stable set of G) is at most k, if and only if the polynomial f_G vanishes whenever $k+1$ variables are equal. For $0 \leqslant k < n$, let $I(k+1,n)$ denote the ideal of the ring $\mathbb{Z}[x_1,\dots,x_n]$ consisting of all polynomials which vanish whenever $k+1$ variables are equal. Hence, $f_G \in I(k+1,n)$ if and only if $c(G) \leqslant k$. Li and Li (1981) proved the following "Nullstellensatz"-type result, which supplies a set of generators of $I(k+1,n)$. In view of the preceding remark, this theorem supplies a characterization (though, maybe, not a very convenient one) for all graphs G whose independence number is at most k.

Theorem 7.1. *Put $V=\{1,2,\dots,n\}$ and let C denote the set of all graphs H on V that consist of k node-disjoint complete graphs whose cardinalities are as equal as possible. Then $\{f_H \colon H \in C\}$ is a set of generators of $I(k+1,n)$. In particular, a graph G has an independence number at most k if and only if there are polynomials $\{g_H \colon H \in C)$ such that $f_G = \sum \{g_H \cdot f_H \colon H \in C\}$.*

Kleitman and Lovász proved a similar result for graphs whose chromatic number is at least k. They showed that a graph G has a chromatic number at least k if and only if f_G belongs to the ideal generated by the polynomials of complete graphs of k nodes from V. Another result of this type appears in Alon and Tarsi (1992); the chromatic number of G is at least k if and only if f_G lies in the ideal generated by the polynomials $x_i^{k-1}-1$.

It is worth noting that, as is well known, the decision problem "Given a graph G and an integer k, is the independence number of G at most k?" as well as the corresponding problem of coloring, are both coNP-complete, and hence it is not reasonable to expect to find a completely satisfactory characterization of the corresponding sets of graphs.

Li and Li (1981) show how to apply Theorem 7.1 to deduce Turán's theorem, which states that the minimum possible number of edges of a graph G on n nodes, whose independence number is at most k, is the number of edges of a node disjoint union of k complete graphs of total order n whose cardinalities are as equal as possible. Indeed, since f_G belongs to the ideal $I(k+1,n)$ which is generated by the graph polynomials $\{f_H\colon H \in C\}$, the degree of f_G, which is precisely the number of its edges, is at least the minimum degree of a generator f_H. Here all generators have the same degree and Turán's theorem follows.

Another combinatorial result whose proof is related to Hilbert's Nullstellensatz deals with the problem of balancing sets of vectors. For an even integer n, let $K(n)$ denote the minimum k for which there exist ± 1 vectors $v_1, v_2, \ldots, v_k$ of dimension n such that for any ± 1 vector w of dimension n, there is an i, $1 \leqslant i \leqslant k$, such that $v_i \cdot w = 0$, i.e., v_i is orthogonal to w. Motivated by a problem in data communication, Knuth showed that $K(n) \leqslant n$ by the following simple construction. For $0 \leqslant i \leqslant n$, let v_i be a vector of i -1 entries followed by $n-i$ 1 entries. We claim that for any ± 1 vector w of dimension n, $w \cdot v_i = 0$ for some $1 \leqslant i \leqslant n$. To see this, note that $w \cdot v_0 = -w \cdot v_n$ while $w \cdot v_i = w \cdot v_{i+1} \pm 2$ for each $i < n$. Since $w \cdot v_j \equiv 0 \pmod 2$ for all i, an obvious "discrete intermediate value" theorem implies that $w \cdot v_i = 0$ for some i, $1 \leqslant i \leqslant n$, as claimed.

As shown by Alon et al. (1988), this construction is optimal, i.e., $K(n) = n$ for all even n. Let us sketch the proof of the lower bound. For simplicity, we consider only the case $n \equiv 0 \pmod 4$. Let U be the set of all ± 1 vectors of dimension n. A vector $u \in U$ is *even* if it has an even number of -1 entries, otherwise it is odd. Let $V \subset U$ be a set of vectors such that for every $u \in U$ there is a $v \in V$ with $v \cdot u = 0$. We must show that $|V| \geqslant n$. Let V_0 be the set of all even vectors of V and let V_1 be the set of all odd vectors of V. Consider the following polynomial in $y = (y_1, \ldots, y_n)$:

$$P(y) = \prod_{v \in V_0} (v \cdot y) .$$

Since $n \equiv 0 \pmod 4$, $v_1 \cdot v_2 \equiv 0 \pmod 2$ for all $v_1, v_2 \in U$. Also, one can easily check that for every $v_1, v_2 \in U$, $v_1 \cdot v_2 = 0 \pmod 4$ if and only if both v_1 and v_2 are even or both are odd. Otherwise $v_1 \cdot v_2 \equiv 2 \pmod 4$. Therefore, for every even $y \in U$, $P(y) = 0$, whereas for every odd $y \in U$, $P(y) \neq 0$. Hence $P(y)$ vanishes on the zero set of the ideal generated by $y_1^2 - 1, y_2^2 - 1, \ldots, y_n^2 - 1, y_1 y_2 \cdots y_n - 1$. By Hilbert's Nullstellensatz, a power of P belongs to this ideal. It is not too difficult (but a little tedious) to show that this implies that if the degree of P is less than $\frac{1}{2}n$ then it vanishes identically, contradicting the fact that $P(y) \neq 0$ for every odd $y \in U$. Thus $\deg P = |V_0| \geqslant \frac{1}{2}n$. Similarly $|V_1| \geqslant \frac{1}{2}n$ and hence $|V| \geqslant n$, completing the proof that $k(n) = n$. A more elementary proof of a somewhat more general result appears in the above mentioned paper.

7.2. Rédei's theorems on lacunary polynomials over finite fields

Let $f(x) = \sum_{i=0}^{n} a_i x^i$ be a polynomial over a finite field $\mathrm{GF}(q)$. It is called *lacunary* if it is 0 or a monomial, or if there are j, k satisfying $a_j \neq 0$, $a_k \neq 0$, $j+2 \leqslant k$ and $a_{j+1} = \cdots = a_{k-1} = 0$. It is called *fully reducible* if it is a product of linear factors over $\mathrm{GF}(q)$.

Rédei (1973) developed a theory which enabled him to give a complete characterization of certain fully reducible lacunary polynomials over finite fields. The main part of this characterization is a determination of all fully reducible polynomials $f(x) = \sum_{i=0}^{q} a_i x^i$ over $\mathrm{GF}(q)$ that satisfy $f'(x) \neq 0$, $a_q \neq 0$ and $a_i = 0$ for $(q+1)/2 < i < q$. Although the full statement of Rédei's theorem is somewhat complicated we give it here, as it seems to be important and yet little known.

Theorem 7.2. *Let $f(x) = \sum_{i=0}^{q} a_i x^i$ be a fully reducible polynomial over* $\mathrm{GF}(q)$, *where $q = p^n$ for a prime $p > 7$, and suppose that $a_q = 1$, $a_i = 0$ for $(q+1)/2 < i < q$ and $f'(x) \neq 0$. Suppose, further, that $f(x) \neq x^q - x$. Then $a_{(q+1)/2} \neq 0$ and $f(x)$ can be obtained as follows. Let σ be $+1$ or -1 and let $p = p_0 < p_1 < \cdots < p_k = q$ be integers satisfying $p_0 | p_1 | \cdots | p_k$ and $(p_0 - 1) | (p_1 - 1) | \cdots | (p_k - 1)$. Let $a_0, a_1, \ldots, a_{k-1}$ be elements of* $\mathrm{GF}(p)$ *satisfying $\chi(a_i) \in \{0, \sigma\}$ for $1 \leqslant i < k$, where χ is the quadratic character, and suppose the elements $p_i \in$* $\mathrm{GF}(p_i)$ *are defined, for $0 \leqslant i < k$, by*

$$\rho_0 = a_0, \qquad \rho_1 = (a_1 - \rho_0)^{(p_0-1)/(p_1-1)},$$

$$\rho_2 = \left((a_2 - \rho_0)^{(p_0-1)/(p_1-1)} - \rho_1\right)^{(p_1-1)/(p_2-1)}, \ldots,$$

$$\rho_i = \Big(\Big(\ldots\Big((a_i - \rho_0)^{(p_0-1)/(p_1-1)} - \rho_1\Big)^{(p_1-1)/(p_2-1)}$$

$$- \cdots - \rho_{i-2}\Big)^{(p_{i-2}-1)/(p_{i-1}-1)} - \rho_{i-1}\Big)^{(p_{i-1}-1)/(p_i-1)}$$

and $p_k \in$ $\mathrm{GF}(p_k)$ *is arbitrary. Define*

$$c(x) = \Big(\ldots\Big(\big(x + \rho_k\big)^{(p_k-1)/(p_{k-1}-1)} + \rho_{k-1}\Big)\ldots\Big)^{(p_1-1)/(p_0-1)} + \rho_0$$

and choose $\tau \in \{0, 1\}$.

Then

$$f(x) = \frac{x^q - x}{c(x)^{(p-1)/2} + \sigma}\left(c(x)^{(p-1)/2} - \sigma\tau\right).$$

Although this theorem may look too complicated to apply, Rédei gave many highly nontrivial, fascinating applications of it to several problems in number theory, group theory and combinatorics. Lovász and Schrijver (1981) gave a short proof to some of these applications. Amazingly, the basic idea in this proof is just the simple fact that every function over a finite field is a polynomial. This enables one to derive combinatorial results by manipulating with these polynomials. Using this idea, Lovász and Schrijver give a short proof of the following result of Rédei.

Theorem 7.3. *For a prime p, any set X of p points, not all on a line, in the affine plane* $\mathrm{AG}(2, p)$, *determines at least $(p+3)/2$ directions. (X determines a direction if there is a line in this direction containing at least two points of X.)*

Blokhuis and Seidel (1985) showed that Wielandt's visibility theorem is an almost direct consequence of this result. It also has some applications in group factoring. Let G be a finite abelian group, written additively, and suppose $A_1, A_2, \ldots, A_m$ are subsets of G, each containing 0. We say that G has an $(A_1, \ldots, A_m)$ factoring and write $G = (A_1, A_2, \ldots, A_m)$ if every element of G is uniquely expressible as a sum $a_1 + a_2 + \cdots + a_m$, where $a_i \in A_i$. Using Theorem 7.3, one can show that if $G \simeq \mathbb{Z}_p \oplus \mathbb{Z}_p$, where p is a prime and $G = (A, B)$, then either A or B is a subgroup. Indeed, G is naturally isomorphic to AG$(2, p)$. If $|A|, |B| > 1$ then $|A| = |B| = p$. It is not too difficult to check (see Lovász and Schrijver 1981) that no direction is determined by both A and B. Hence either A or B determines at most half of the directions, i.e., less than $(p+3)/2$ directions. By Theorem 7.3 this set is a line, and since it contains 0, it is a subgroup.

Rédei obtained a far reaching generalization of this result. Using group characters, an appropriate factorization of polynomials and the group ring of G over the integers, he proved the following.

Theorem 7.4 (Rédei 1965). *If G is a finite abelian group that has an $(A_1, \ldots, A_m)$ factoring, where each A_i has a prime order, then at least one A_i is a subgroup.*

This theorem generalizes Hajós theorem, (cf. Fuchs 1967) which is probably the most dramatic work in factoring groups, and which solved a tiling problem raised by Minkowski in 1907.

Theorem 7.3 is a special case of a more general result of Rédei (1973, pp. 225, 226), which asserts that the number of directions m determined by a set of q points, not all on a line, in the affine plane AG$(2, q)$, where $q = p^n$ is a prime power, is at least

$$m \geqslant \frac{q-1}{p^{\lfloor n/2 \rfloor} + 1} + 1 . \tag{7.1}$$

Blokhuis and Brouwer (1986) found a nice way to combine this result with the Jamison–Brouwer–Schrijver theorem (see Theorem 6.5), and derive a bound for the size of non-trivial blocking sets in Desarguesian projective planes. Let PG$(2, q)$ denote the projective plane over GF(q). A *blocking set* in PG$(2, q)$ is a set that intersects every line. It is *nontrivial* if it contains no line. Bruen showed that any non-trivial blocking set S in PG$(2, q)$ contains at least $q + \sqrt{q} + 1$ points, and equality holds iff q is a square and S is the set of points of a Baer subplane. He also noticed, together with Thas, that there is a connection between Rédei's results and blocking sets. This connection was later applied by Blokhuis and Brouwer to prove that if $q = p^n > 7$ is a non-square, odd prime power, $q \neq 27$, then any nontrivial blocking set in PG$(2, q)$ contains more than $q + \sqrt{2q}$ points. The proof is very short; let S be a minimal, nontrivial blocking set, $|S| = q + k$. If there exists a line containing k of these points, then make it the line at infinity. The remaining q points now block all the lines of the affine plane except in k directions, and hence they determine at most these k directions. By Rédei's result stated in (7.1),

$$k \geqslant \frac{q-1}{p^{\lfloor n/2 \rfloor} + 1} + 1 > \sqrt{2q} ,$$

as needed. Thus, we may assume that no line contains more than $(k-1)$ points of S. Let v be an arbitrary point of S. By the minimality of S, there exists a tangent through v (i.e., a line containing only this point from S). Make this line the line at infinity and observe that the remaining points cover all lines of the affine plane except the other tangents at this point. By Theorem 6.5, there must be at least $q-k$ such other tangents. Thus, there are at least $q-k+1$ tangents through any point of S. Since there are no k points of S on a line, there are at most $q-1$ tangents through any point not in S. Hence, by counting the incident pairs of the form (tangent, point not in S) we conclude that

$$(q+k)(q-k+1)q \leqslant (q^2-k+1)(q-1)\ ,$$

which gives $k > \sqrt{2q}$, completing the proof.

For the (odd) prime case $q=p$, the above estimate has recently been improved considerably in an elegant paper of Blokhuis (1994) to $3(p+1)/2$, which is optimal.

7.3. *Hilbert's basis theorem and Ehrenfeucht's conjecture in language theory*

For a (finite) alphabet A, let A^* denote, as usual, the set of all finite words over A. For two alphabets A, B, a function $f : A^* \to B^*$ is a *morphism* if for every $x, y \in A^*$, $f(xy) = f(x)f(y)$, where xy and $f(x)f(y)$ denote here the concatination of x, y and that of $f(x), f(y)$, respectively.

Let A be a finite alphabet and let $\mathscr{L} \subset A^*$ be an arbitrary language over A. Ehrenfeucht conjectured that there is always a finite set $F \subset \mathscr{L}$ such that for any alphabet B and for any two morphisms $g, h : A^* \to B^*$, $g(x) = h(x)$ for all $x \in \mathscr{L}$ if and only if $g(x) = h(x)$ for all $x \in F$. We call such an F a *test set* for $\mathscr{L}$.

This conjecture has been solved, independently, by Albert and Lawrence, by McNaughton and by V.S. Guba (cf. Salomaa 1985). All proofs reduce the conjecture to Hilbert's basis theorem, which is the following.

Theorem 7.5. *Every ideal in the polynomial ring $\mathbb{Z}[x_1, \ldots, x_n]$ is finitely generated. Hence any infinite system S of polynomial equations over $\mathbb{Z}$ is equivalent to some finite subsystem S' of it (i.e., S and S' have the same solutions in the complex field).*

Hilbert's basis theorem can be proved by a rather simple induction on n (see, e.g., Van der Waerden 1931). A special case of it plays an important role in integer programming, see chapter 30.

Ehrenfeucht's conjecture is reduced to Hilbert's theorem in two steps, as outlined below.

Step 1. A system of equations $W^{(i)} = \overline{W}^{(i)}$ $(i \in I)$ where $W^{(i)}$ and $\overline{W}^{(i)}$ are words in C^*, has a *solution* f if there exists an alphabet D and a morphism $f : C^* \to D^*$ such that $f(W^{(i)}) = f(\overline{W}^{(i)})$ for all $i \in I$. Two systems of word equations are *equivalent* if they have the same solutions. It is not too difficult to reduce Ehrenfeucht's conjecture to the following statement about word equations.

Statement 7.6. *Every system of word equations is equivalent to a finite subsystem of it.*

Step 2. Statement 7.6 is reduced to Theorem 7.5 by constructing, for any system E of word equations over an alphabet C, a system S of polynomials such that every solution of E corresponds to a solution of a certain type of S. (The system S might have some other solutions, as well.)

Since Step 2 is the crucial part of the proof let us briefly describe it. The basic idea is the following. If the alphabet D has n letters, then any word in D^* corresponds, naturally, to the number it represents in base n. If $f : C^* \to D^*$ is a morphism, then for every word $W \in C^*$, $f(W)$, considered as the number it describes, can be expressed as a polynomial in the numbers $f(c)$ for $c \in C$ and the numbers $n^{\text{length}(f(c))}$, where length $(f(c))$ is the number of letters in the word $f(c)$. Therefore, by introducing variables for the $2|c|$ numbers $f(c)$ and $n^{\text{length}(f(c))}$ for $c \in C$, we can replace each word equation by two polynomial equations. Being more precise now, let us introduce, for each letter $c \in C$, two variables c_1 and c_2. (We will later substitute $f(c)$ for c_1 and $n^{\text{length}(f(c))}$ for c_2.) For any word $W = c^1c^2 \cdots c^k \in C^*$ define

$$P_1(W) = c_1^1c_2^2c_2^3 \cdots c_2^k + c_1^2 \cdot c_2^3 \cdot c_2^4 \cdots c_2^k + \cdots + c_1^{k-1}c_2^k + c_1^k$$

and

$$P_2(W) = c_2^1c_2^2 \cdots c_2^k \,. \tag{7.2}$$

Also, for the empty word λ, $P_1(\lambda) = 0$ and $P_2(\lambda) = 1$. Given the system E of word equations $W^{(i)} = \overline{W}^{(i)}$ $(i \in I)$, let S be the system of polynomial equations $P_1(W^{(i)}) = P_1(\overline{W}^{(i)})$ and $P_2(W^{(i)}) = P_2(\overline{W}^{(i)})$ $(i \in I)$. By construction, for every alphabet D of n letters and every morphism $f : C^* \to D^*$, f is a solution of E if and only if $c_1 = f(c)$ and $c_2 = n^{\text{length}(f(c))}$ $(c \in C)$ is a solution of S. Therefore, the existence of a finite subsystem of S equivalent to it, which follows from Theorem 7.5, supplies the existence of a finite subsystem of E equivalent to E. For more details, including the (simple) proof of the equivalence between Ehrenfeucht's conjecture and Statement 7.6, see Salomaa (1985).

We note that the decision problem: "Given a (recursively enumerable) language $\mathscr{L} \subset A^*$ and two morphisms $g,h : A^* \to B^*$, is $g(x) = h(x)$ for all $x \in \mathscr{L}$?" is undecidable, and thus there is no "constructive" proof of Ehrenfeucht's conjecture (i.e., a proof that actually produces a finite test set for $\mathscr{L}$ from its description).

8. Hyperbolic geometry and triangulations of polytopes and polygons

Let P be a 3-dimensional simplicial polytope. Let $T(P)$ denote the minimum number of tetrahedra, each being the convex hull of four vertices of P, whose union covers P. For $n \geqslant 4$, let $T(n)$ be $\max T(P)$, where the maximum is taken over all simplicial polytopes P with n vertices.

It is easy to check that for every $n > 12$, $T(n) \leqslant 2n - 10$. Indeed, a simplicial 3-polytope P on n vertices has $2n - 4$ faces and $3n - 6$ edges. If $n > 12$, there is a vertex v of P incident with at least 6 faces. For each other face f of P, let S_f be a tetrahedron whose vertices are v and the three vertices of f. These tetrahedra cover P, and their number is at most $2n-10$.

Sleator et al. (1986) proved that $T(n) \geqslant 2n-10$ (and hence equals $2n-10$) for infinitely many values of n. Their interesting proof uses hyperbolic geometry. Here is an outline of the idea. If one can construct a polytope P and show, somehow, that the volume of each tetrahedron on 4 of its vertices is at most a fraction $1/\ell$ of the volume of P, then the inequality $T(P) \geqslant \ell$ follows. Unfortunately, the largest ℓ for which the previous statement holds is a constant, independent of the number of vertices of P. Thus, instead of using the usual Euclidean space $\mathbb{R}^3$, we embed P in the 3-dimensional hyperbolic space (and observe that any cover of P by tetrahedra in $\mathbb{R}^3$ corresponds to a cover of the same size of P here). In this new space, the volume of each tetrahedron is bounded by a constant C_0, and thus we need only construct a polytope P whose volume is at least $\ell \cdot C_0$. For $\ell = 2n - O(\sqrt{n})$ a construction of such a polytope on n vertices is not too difficult. The reader is referred to Coxeter (1956) for the fundamentals of hyperbolic geometry. The 3-dimensional hyperbolic space can be viewed as an upper half space whose boundary is the complex plane, plus a point denoted ∞. A geodesic here is a semicircle perpendicular to the complex plane, or a line perpendicular to this plane, that goes to ∞. Any tetrahedron whose base forms an equilateral triangle on the complex plane and whose fourth vertex is ∞ is a tetrahedron of maximum volume. Consider a tessellation of the complex plane by equilateral triangles, and let S be a set of $6k^2$ such triangles whose union is hexagonal, with k edges on each side. This hexagon has $3k^2+3k+1$ vertices. Let P be the polytope whose vertices are ∞ and these vertices. Since P is the union of $6k^2$ tetrahedra of maximal volume, its volume is $6k^2 \cdot C_0$. This shows that $T(3k^2+3k+2) \geqslant 6k^2$, and hence that $T(n) \geqslant 2n - O(\sqrt{n})$. For the sharper estimate $T(n) \geqslant 2n-10$, see Sleator et al. (1986).

The problem of covering a polytope by tetrahedra is related to another interesting combinatorial problem. Let G be a labeled convex polygon with n vertices in the plane, and consider a planar triangulation of G with no interior vertices. We call the n sides of G *edges* and the chords that divide it into triangles are called *diagonals*.

A *diagonal flip* is an operation that transforms one triangulation of G into another by removing a diagonal, thus creating a face with four sides, and by inserting the opposite diagonal of this resulting quadrilateral. The *distance* $d(\tau_1, \tau_2)$ between two triangulations τ_1 and τ_2 of G is the minimum number of diagonal flips needed to transform one into the other. Motivated by a data-structure problem on dynamic trees, Sleator et al. (1986) considered the problem of determining or estimating $d(n) = \max d(\tau_1, \tau_2)$, where τ_1 and τ_2 range over all triangulations of a labeled n-gon. It is easy to see that $d(n) \leqslant 2n-10$ for all $n > 12$. Somewhat surprisingly, a lower bound for $d(n)$, showing that $d(n) = 2n-10$ for infinitely many values of n, can be extracted from the corresponding result for $T(n)$ – the maximum value of the minimum number of tetrahedra needed to cover a convex n-polytope. Here is an outline of the idea.

Let P be a convex simplicial n-polytope whose graph is Hamiltonian such that

$T(P)$ is as large as possible. (By the Sleator–Tarjan–Thurston result, there is such P with $T(P) = 2n - 10$ for infinitely many values of n.) Cut P along the edges of the Hamilton cycle to obtain two triangulated parts. Denote these two triangulations by τ_1 and τ_2. We claim that $d(\tau_1, \tau_2) \geqslant T(P) = 2n - 10$ (and hence $d(\tau_1, \tau_2) = 2n - 10$). To see this we show that P can be covered by $d(\tau_1, \tau_2)$ tetrahedra. Consider a sequence of $d(\tau_1, \tau_2)$ diagonal flips that transform τ_1 into τ_2. Imagine a planar base with triangulation τ_1 drawn on it. Suppose the first diagonal flip replaces the diagonal (a, c) with the diagonal (b, d). Create a flat quadrilateral with the same shape as (a, b, c, d). On its back draw the diagonal (a, c) and on its front draw the diagonal (b, d). Now place this quadrilateral onto the base in the appropriate place, with the diagonal (a, c) down and (b, d) up. Looking from the top we see a picture of the triangulation obtained from τ_1 by making the first diagonal flip. For each successive move we create an additional quadrilateral and place it onto the base. After placing $d(\tau_1, \tau_2)$ such quadrilaterals we will see τ_2 when we view the base from the top. We can now inflate each quadrilateral slightly, to make it into a tetrahedron. The resulting stack of quadrilaterals forms a covering of P by $d(\tau_1, \tau_2)$ tetrahedra, as needed. For more details and several other related results the reader is referred to Sleator et al. (1986).

9. The Erdős–Moser conjecture and the hard Lefschetz theorem

For a finite subset S of $\mathbb{R}$, and for $k \in \mathbb{R}$, let $f(S, k)$ denote the number of subsets of S whose elements sum to k. Erdős and Moser conjectured, in 1965, that for every set S of $2n + 1$ distinct real numbers, and any k,

$$f(S, k) \leqslant f(\{-n, -n+1, \ldots, n\}, 0) . \tag{9.1}$$

Similarly, it was conjectured that for every set T of n distinct positive numbers and any k

$$f(T, k) \leqslant f(\{1, 2, \ldots, n\}, [n(n+1)/4]) . \tag{9.2}$$

Both (9.1) and (9.2) follow from the results of Stanley (1980) (see also Stanley 1983). Surprisingly, Stanley's results depend on some deep results from algebraic geometry and in particular on the hard Lefschetz theorem, stated in chapter 34. A somewhat more elementary, similar proof was given later, by Proctor (1982), whose proof involved representations of the Lie algebra $s\ell(2, C)$. However, there is no known purely combinatorial proof.

To prove (9.2) it is useful to define the following partially ordered set $M(n)$. The elements of $M(n)$ are all ordered sets of integers $(a_1, a_2, \ldots, a_k)$ where $n \geqslant a_1 > a_2 > \cdots > a_k \geqslant 1$, and $(a_1, \ldots, a_k) \geqslant (b_1, \ldots, b_j)$ if $k \geqslant j$ and $a_1 \geqslant b_1, \ldots, a_j \geqslant b_j$. Put $M(n)_r = \{(a_1, \ldots, a_k) \in M(n)\colon \sum_{i=1}^{k} a_i = r\}$ and notice that $|M(n)_r| = f(\{1, 2, \ldots, n\}, r)$. Define also $N = \binom{n+1}{2}$. An easy lemma, first observed by Lindström, states that if $M(n)_{[N/2]}$ is the biggest antichain of $M(n)$, then (9.2) holds. Stanley proved that $M(n)_{[N/2]}$ is the biggest antichain of $M(n)$ by showing that for every $0 \leqslant i \leqslant [N/2]$ there exist $M(n)_i$ pairwise disjoint chains $x_i < x_{i+1} < \cdots < x_{N-i}$ in $M(n)$, where $x_j \in M(n)_j$. The proof uses the linear algebra method, whose many applications in combinatorics are described in chapter 31. However, the construction of the necessary linear mappings is highly nontrivial. We construct linear transformations $\varphi_i : V_i \rightarrow V_{i+1}$ for $0 \leqslant i < N$, where V_i is the complex vector space with basis $M(n)_i$, such that for $0 \leqslant i \leqslant [N/2]$, $\varphi_{N-i-1} \circ \varphi_{N-i-2} \circ \cdots \circ \varphi_i : V_i \rightarrow V_{N-i}$ is invertible and for $x \in M(n)_i$ and $\varphi_i(x) = \sum \{c_y \cdot y : y \in M(n)_{i+1}\}$, $c_y \neq 0$ implies $y > x$. This, in turn, supplies the existence of the desired pairwise disjoint chains in $M(n)$.

The existence of these mappings is established using the hard Lefschetz theorem, stated in chapter 34. For more details and more general results see Stanley (1980). Several other fascinating combinatorial applications of the hard Lefschetz theorem appear in Stanley (1983) and some of its references.

References

Ajtai, M., J. Komlós and E. Szemerédi

[1983] Sorting in $c \log n$ parallel steps, *Combinatorica* **3**, 1–19.

Alon, N.

[1986a] Eigenvalues and expanders, *Combinatorica* **6**, 83–96.

[1986b] The number of polytopes, configurations and real matroids, *Mathematika* **33**, 62–71.

Alon, N., and V.D. Milman

[1984] Eigenvalues, expanders and superconcentrators, in: *Proc. 25th Annu. Symp. on Foundations of Computer Science, Florida* (IEEE Computer Society Press, New York) pp. 320–322.

[1985] λ_1, isoperimetric inequalities for graphs and superconcentrators, *J. Combin. Theory B* **38**, 73–88.

Alon, N., and E.R. Scheinerman

[1988] Degrees of freedom versus dimension for containment orders, *Order* **5**, 11–16.

Alon, N., and M. Tarsi

[1992] Colorings and orientations of graphs, *Combinatorica* **12**, 125–134.

Alon, N., S. Friedland and G. Kalai

[1984] Regular subgraphs of almost regular graphs, *J. Combin. Theory B* **37**, 79–91.

Alon, N., P. Frankl and V. Rödl

[1985] Geometrical realization of set systems and probabilistic communication complexity, in: *Proc. 26th Annu. Symp. on Foundations of Computer Science, Oregon* (IEEE Computer Society Press, New York) pp. 277–280.

Alon, N., Z. Galil and V.D. Milman

[1987] Better expanders and superconcentrators, *J. Algorithms* **8**, 337–347.

Alon, N., E.E. Bergmann, D. Coppersmith and A.M. Odlyzko

[1988] Balancing sets of vectors, *IEEE Trans. Inform. Theory* **IT-34**, 128–130.

Angluin, D.

[1979] A note on a construction of Margulis, *Inform. Process. Lett.* **8**, 17–19.

Baker, R.C., and W. Schmidt

[1980] Diophantine problems in variables restricted to the values 0 and 1, *J. Number Theory* **12**, 460–486.

Bassalygo, L.A.

[1981] Asymptotically optimal switching circuits, *Problemy Peredachi Informatsii* **17**(3), 81–88 [*Problems Inform. Transmission* **17**(3), 206–211].

Ben-Or, M.

[1983] Lower bounds for algebraic computation trees, in: *Proc. 15th ACM Symp. on the Theory of Computing* (ACM, New York) pp. 80–86.

Biggs, N.L.

[1989] Cubic graphs with large girth, in: *Ann. N.Y. Acad. Sci.* **555**, eds. G.S. Blum, R.L. Graham and J. Malkevitch, pp. 56–62.

Biggs, N.L., and M.H. Hoare

[1983] The sextet construction for cubic graphs, *Combinatorica* **3**, 153–165.

Björner, A., L. Lovász and A.C.C. Yao

[1992] Linear decision trees: Volume estimates and topological bounds, in: *Proc. 24th ACM Symp. on the Theory of Computing* (ACM, New York) pp. 170–177.

Blokhuis, A.

[1994] On the size of a blocking set in $PG(2,p)$, *Combinatorica* **14**, 111–114.

Blokhuis, A., and A.E. Brouwer

[1986] Blocking sets in desarguesian projective planes, *Bull. London Math. Soc.* **18**, 132–134.

Blokhuis, A., and J.J. Seidel

[1985] Remark on Wielandt's visibility theorem, *Linear Algebra Appl.* **71**, 29–30.

Bollobás, B.

[1978] *Extremal Graph Theory* (Academic Press, New York).

Bollobás, B., and A.G. Thomason

[1981] Graphs which contain all small graphs, *European J. Combin.* **2**, 13–15.

Borevich, Z.I., and I.R. Shafarevich

[1966] *Number Theory* (Academic Press, New York).

Brouwer, A.E., and A. Schrijver

[1978] The blocking number of an affine space, *J. Combin. Theory A* **24**, 251–253.

Buck, M.W.

[1986] Expanders and diffusers, *SIAM J. Algebraic Discrete Methods* **7**, 282–304.

Chung, F.R.K.

[1978] On concentrators, superconcentrators, generalizers and nonblocking networks, *Bell. Sys. Tech. J.* **58**, 1765–1777.

Coxeter, H.S.M.

[1956] *Non-Euclidean Geometry* (University of Toronto Press, Toronto).

Erdős, P.

[1963] On a problem in graph theory, *Math. Gaz.* **47**, 220–223.

Erdős, P., A. Ginzburg and A. Ziv

[1961] Theorem in the additive number theory, *Bull. Res. Council Israel* **10F**, 41–43.

Frankl, P., and R.M. Wilson

[1981] Intersection theorems with geometric consequences, *Combinatorica* **1**, 357–368.

Fuchs, L.

[1967] *Abelian Groups* (Pergamon Press, Oxford).

Gabber, O., and Z. Galil

[1981] Explicit construction of linear sized superconcentrators, *J. Comput. Sys. Sci.* **22**, 407–420.

Goodman, J.E., and R. Pollack

[1986] Upper bounds for configurations and polytopes in $\mathbb{R}^d$, *Discrete Comput. Geom.* **1**, 219–227.

Graham, R.L., and J.H. Spencer

[1971] A constructive solution to a tournament problem, *Canad. Math. Bull.* **14**, 45–48.

Grünbaum, B.

[1967] *Convex Polytopes* (Wiley Interscience, London).

Hocking, J.G., and G.S. Young

[1961] *Topology* (Addison-Wesley, Reading, MA).

Imrich, W.

[1984] Explicit construction of regular graphs without small cycles, *Combinatorica* **4**, 53–59.

Jamison, R.E.

[1977] Covering finite fields with cosets of subspaces, *J. Combin. Theory A* **22**, 253–266.

Jimbo, Sh., and A. Maruoka

[1985] Expanders obtained from affine transformations, in: *Proc. 17th Annu. ACM Symp. on Theory of Computing* (ACM, New York) pp. 88–97.

Kazhdan, D.

[1967] Connection of the dual space of a group with the structure of its closed subgroups, *Funct. Anal. Appl.* **1**, 63–65.

Li, S.-Y.R., and W.-C.W. Li

[1981] Independence number of graphs and generators of ideals, *Combinatorica* **1**, 55–61.

Lovász, L., and A. Schrijver

[1981] Remarks on a theorem of Rédei, *Studia Sci. Math. Hungar* **16**, 449–454.

Lubotzky, A., R. Phillips and P. Sarnak

[1986] Explicit expanders and the Ramanujan conjectures, in: *Proc. 18th Annu. ACM Symp. on Theory of Computing* (ACM, New York) pp. 240–246.

[1988] Ramanujan graphs, *Combinatorica* **8**(3), 261–277.

Magnus, W., A. Karrass and D. Solitar

[1966] *Combinatorial Group Theory* (Interscience, New York).

Margulis, G.A.

[1973] Explicit constructions of concentrators, *Problemy Peredachi Informatsii* **9**(4), 71–80 [*Problems Inform. Transmission* 9(4), 325–332].

[1982] Graphs without short cycles, *Combinatorica* **2**, 71–78.

[1984] Arithmetic groups and graphs without short cycles, in: *Proc. 6th Int. Symp. on Information Theory, Tashkent,* Vol. 1, pp. 123–125 (in Russian).

[1988] Explicit group theoretic constructions of combinatorial schemes and their applications for the construction of expanders and concentrators, *Problemy Peredachi Informatsii* **24**, 51–60 [*Problems Inform. Transmission* **24**, 39–46].

Milnor, J.

[1964] On the Betti numbers of real varieties, *Proc. Amer. Math. Soc.* **15**, 275–280.

Newman, M.

[1972] *Integral Matrices* (Academic Press, New York).

Oleĭnik, O.A., and I.B. Petrovskiĭ

[1962] On the topology of real algebraic surfaces, *Izv. Akad. Nauk SSSR* **13**, 389–402 [*Trans. Amer. Math. Soc. (1)* 7, 399–417].

Olson, J.E.

[1969a] A combinatorial problem on finite abelian groups, *J. Number Theory* **1**, 8–10.

[1969b] A combinatorial problem on finite abelian groups II, *J. Number Theory* 1, 195–199.

Paturi, R., and J. Simon

[1984] Probabilistic communication complexity, in: *Proc. 25th Annu. Symp. on Foundations of Computer Science, Florida* (IEEE Computer Society Press, New York) pp. 118–126.

Paul, W.J., R.E. Tarjan and J.R. Celoni

[1977] Space bounds for a game on graphs, *Math. Sys. Theory* **10**, 239–251.

Pinsker, M.

[1973] On the complexity of a concentrator, in: *7th Int. Teletraffic Conf. Stockholm* 318/1–318/4.

Pippenger, N.

[1977] Superconcentrators, *SIAM J. Comput.* **6**, 298–304.

Proctor, R.

[1982] Representations of $sl(2,\mathbb{C})$ on posets and the Sperner property, *SIAM J. Algebraic Discrete Methods* **3**, 275–280.

Pyber, L.

[1985] Regular subgraphs of dense graphs, *Combinatorica* **5**, 347–349.

Ramanujan, S.

[1916] On certain arithmetical functions, *Trans. Cambridge Philos. Soc.* **22**, 159–184.

Rédei, L.

[1965] Die neue Theorie der endlichen abelschen Gruppen und Verallgemeinerung des Hauptsatzes von Hajós, *Acta Math. Acad. Sci. Hung.* **16**, 329–373.

[1973] *Lacunary Polynomials Over Finite Fields* (Elsevier, North-Holland, Amsterdam).

Salomaa, A.

[1985] The Ehrenfeucht conjecture: a proof for language theorists, *Bull. European Assoc. Theor. Comput. Sci.* **27**, 71–82.

Schmidt, K.

[1980] Asymptotically invariant sequences and an action of $S\ell(2,\mathbb{Z})$ on the 2-sphere, *Israel J. Math.* **37**, 193–208.

Schmidt, W.M.

[1976] *Equations over Finite Fields. An Elementary Approach. Lecture Notes in Mathematics*, Vol. 536 (Springer, Berlin).

Sleator, D.D., R.E. Trajan and W.P. Thurston

[1986] Rotation distance, triangulations and hyperbolic geometry, in: *Proc. 18th Annu. ACM Symp. on Theory of Computing* (ACM, New York) pp. 122 135.

Stanley, R.P.

[1980] Weyl groups, the hard Lefschetz theorem, and the Sperner property, *SIAM J. Algebraic Discrete Methods* **1**, 168-184.

[1983] Combinatorial applications of the hard Lefschetz theorem, in: *Proc. Int. Congr of Mathematicians, Warszawa*, pp. 447-453.

Steele, J.M., and A.C. Yao

[1982] Lower bounds for algebraic trees, *J. Algorithms* **3**, 1-8.

Tanner, R.M.

[1984] Explicit construction of concentrators from generalized N-gons, *SIAM J. Algebraic Discrete Methods* **5**, 287–293.

Taškinov, V.A.

[1982] Regular subgraphs of regular graphs, *Soviet Math. Dokl.* **26**, 37–38.

Thom, R.

[1965] Sur l'homologie des variétés algebraiques réelles, in: *Differential and Combinatorial Topology*, ed. S.S. Cairns (Princeton University Press, Princeton, NJ) pp. 255–265.

Tompa, M.

[1980] Time space trade-offs for computing functions using connectivity properties of their circuits, *J. Comput. Sys. Sci.* **20**, 118–132.

Valiant, L.G.

[1976] Graph theoretic properties in computational complexity, *J. Comput. Sys. Sci.* **13**, 278–285.

Van der Waerden, B.L.

[1931] *Modern Algebra II* (Julius Springer, Berlin).

Van Emde Boas, P., and D. Kruyswijk

[1969] *A Combinatorial Problem on Finite Abelian Groups III*, Z.W. 1969–008 (Mathematisch Centrum, Amsterdam).

Warren, H.E.

[1968] Lower bounds for approximation by nonlinear manifolds, *Trans. Amer Math. Soc.* **133**, 167–178.

Weil, A.

[1948] Sur les courbes algebraiques et les variétés qui s'en deduisent, *Actualités Sci. Ind.* **1041** (Herman, Paris).

Weiss, A.

[1984] Girths of bipartite sextet graphs, *Combinatorica* **4**, 241 245.

CHAPTER 33

Probabilistic Methods

Joel SPENCER

Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, Room 518,
New York, NY 10012, USA

Contents

HANDBOOK OF COMBINATORICS
Edited by R. Graham, M. Grötschel and L. Lovász

1. Elementary methods

We examine a methodology for proving the existence of combinatorial configurations having certain desired properties. In its basic form a probability space is created whose elements are configurations. It is shown that the probability that a configuration does not have the desired property is less than unity – generally by showing that it is extremely small. With positive probability the configuration is good. Hence, some good configuration exists. The methodology is best described by example. We often give suboptimal results in order to better illustrate the methodology, only making reference to the best-known results.

My joint books with Paul Erdős (Erdős and Spencer 1974) and Noga Alon (Alon and Spencer 1992) and my previous survey paper (Spencer 1978) and monograph (Spencer 1994) deal also with these topics in some detail.

We begin with a simple probability fact which we shall call the *counting sieve*:

$$\text{If } \sum \Pr[A_i] < 1\,, \quad \text{then } \bigwedge \bar{A}_i \neq \emptyset\,.$$

Here the summation and conjunction range over the same index set. It is surprising what results may be obtained from this simple principle.

The Ramsey function $R(k, t)$ is the minimal n such that if the edges of K_n are colored Red and Blue then there exists either a Red K_k or a Blue K_t (see chapter 25). The lower bound $R(k, t) > n$ thus means that there exists a two-coloring of K_n such that neither type of monochromatic subgraph exists. Erdős (1947) inaugurated the modern use of the probabilistic method with the following lower bound on the diagonal function $R(n, n)$.

Theorem 1.1. *If* $\binom{n}{k}2^{1-\binom{k}{2}} < 1$, *then* $R(k, k) > n$.

Proof. Define a probability space on the set of two-colorings of K_n by assuming that each edge is colored Red with probability $\frac{1}{2}$ (otherwise Blue) and that these probabilities are mutually independent. (We may consider this a *Gedanken experiment* in which a fair coin is thrown to determine the color of each edge.) Call a coloring *Bad* if some k-set is monochromatic, otherwise *Good*. For each set S of k vertices let A_S be the event that S is monochromatic. As S has $\binom{k}{2}$ edges and the corresponding coin flips must all be the same, $\Pr[A_S] = 2^{1-\binom{k}{2}}$. There are $\binom{n}{k}$ such S, so our hypothesis and the counting sieve imply that the event $\wedge \bar{A}_S$ is nonempty. The event $\wedge \bar{A}_S$ means precisely that the coloring is Good. Hence there is a Good coloring. Hence $R(k, k) > n$. □

Nearly all applications of the probabilistic method deal with the asymptotic behavior of combinatorial functions. It is essential to have a good feel for asymptotic calculations (see chapter 22). For example, consider finding that n for which $\binom{n}{k}2^{1-\binom{k}{2}}$ is roughly unity. A "seat of the pants" calculation would be to estimate $\binom{n}{k}$ by n^k and $\binom{k}{2} - 1$ by $k^2/2$, so that the critical n is roughly when $n^k 2^{-k^2/2} = 1$, $n = 2^{k/2}$. This is the first term for the lower bound for $R(k, k)$. The upper bound (which is not probabilistic) for $R(k,k)$ is roughly 4^k. A major, and thus far intractable, problem is to find that c for which $R(k,k)$ is roughly c^k.

A more precise calculation using Stirling's Formula gives

$$R(k,k) > [k/e\sqrt{2}]2^{k/2}(1 + o(1)) .$$

A calculation with $k = 100$ shows $R(100, 100) > 3 \times 10^{15}$. A strong threshold behavior is observed if we decrease n slightly, say $n = 2.5 \times 10^{15}$. A coloring is then Bad with probability less than $\binom{n}{100}2^{1-4950} < 10^{-9}$. A coloring created randomly on 2.5×10^{15} vertices will almost certainly be Good.

The properties of random configurations make for fascinating study in their own right, involving a challenging mix of probabilistic and combinatorial techniques. These are described in chapter 6. While our methodologies overlap to a considerable degree in this survey attention is restricted to use of probabilistic methods in order to prove the existence of certain configurations.

Warning. The counting sieve does not work in reverse. Generally other methods must be used to prove the nonexistence of a configuration. In particular, even the existence of an upper bound on $R(k,k)$ – i.e., a proof of Ramsey's Theorem – does not seem to follow from probabilistic considerations.

Theorem 1.2. *If there exists p, $0 \leqslant p \leqslant 1$, such that*

$$\binom{n}{k}p^{\binom{k}{2}} + \binom{n}{t}(1-p)^{\binom{t}{2}} < 1 ,$$

then $R(k,t) > n$.

Proof. We adjust the probability space used in the proof of Theorem 1.1. The space is again the set of two-colorings of K_n but now each edge is colored Red with probability p. For each k-set S let A_S be the event that all edges of S are colored Red, and for each t-set T let B_T be the event that all edges of T are colored Blue. Our assumption and the counting sieve imply the existence of a two-coloring for which no A_S and no B_T hold. Hence $R(k,t) > n$. □

As an example, when $k = 4$ we may take $p = n^{-2/3}$ and $n = t^{3/2+o(1)}$. The hypothesis of Theorem 1.2 is then satisfied so that $R(4,t) > t^{3/2+o(1)}$. Asymptotic bounds on $R(k,t)$ for fixed k have been examined by the present author (Spencer 1977). We will return to the special case $k = 3$ in section 2.

A *tournament* on n players consists of the results of $\binom{n}{2}$ matches, one between every pair of players, in which there are no draws. A tournament is said to have property S_k if for every set of k players there is some other player who beats them all. Do there exist for arbitrarily large k tournaments with property S_k? An affirmative answer was given by Erdős (1963a). Consider a random tournament on n players in which the outcome of each match is decided by the toss of a fair coin. For each set P of k players let A_P be the event that no player outside of P

beats all players in P. Any $y \not\in P$ has probability 2^{-k} of beating all players in P. These events are mutually independent over y as they involve different games. Hence $\Pr[A_P] = (1-2^{-k})^{n-k}$. By the counting sieve: if $\binom{n}{k}(1-2^{-k})^{n-k} < 1$, then $\wedge \bar{A}_p \neq \emptyset$. But this is precisely the event that the tournament has property S_k. For k fixed $\binom{n}{k}(1-2^{-k})^{n-k}$ approaches zero in n so that for n sufficiently large there exist tournaments with property S_k.

A family $\mathcal{F} = \{S_1, \ldots, S_r\}$ of sets is 2-colorable (see chapter 7) is there exists a two-coloring of the underlying points so that no set is monochromatic. Assume each set S_i has precisely n elements. Erdős (1963b) proved that if the number of sets $r < 2^{n-1}$, then $\mathcal{F}$ is 2-colorable. To show this, consider a random coloring of the underlying points. Let A_i be the event that S_i is monochromatic so that $\Pr[A_i] = 2^{1-n}$. With $r < 2^{n-1}$ the counting sieve gives $\wedge \bar{A}_i \neq \emptyset$. This is precisely the event that no set is monochromatic.

Erdős (1964) defined $m(n)$ as the minimal size of a family of n-sets which is not 2-colorable. [In chapter 7 this is denoted by $m_2(n)$.] It is somewhat surprising, and unusual, that an upper bound to this combinatorial function can also be given by probabilistic means. Essentially, the sets now become random and each coloring gives an event. Let $S_1, \ldots, S_r$ be randomly selected n-sets from $\{1, \ldots, v\}$. (That is, each S_i is selected independently and uniformly from among the $\binom{v}{n}$ n-sets.) Let χ be a two-coloring of $\{1, \ldots, v\}$. Assume first that χ has $v/2$ Red and $v/2$ Blue points. For such a given χ any particular S_i will be monochromatic with probability $2\binom{v/2}{n}/\binom{v}{n}$, as it is monochromatic if and only if it is either a subset of the Red points or a subset of the Blue points. Since the S_i are selected *independently*, the probability that no S_i is monochromatic is $[1 - 2\binom{v/2}{n}/\binom{v}{n}]^r$. Let A_χ be the event that under coloring χ no S_i is monochromatic. For any χ,

$$\Pr[A_\chi] \leq \left[1 - 2\binom{v/2}{n}\bigg/\binom{v}{n}\right]^r ,$$

as it is easy to see that $\Pr[A_\chi]$ is maximized when χ is balanced. There are 2^v colorings. The event $\wedge \bar{A}_\chi$, conjunction over the 2^v colorings, is precisely that $\{S_1, \ldots, S_r\}$ is not 2-colorable. Thus $\wedge \bar{A}_\chi \neq \emptyset$ if and only if there is a family $\mathcal{F}$ of r n-sets on v points which is not 2-colorable. Applying the counting sieve: if, for any v, $2^v[1 - 2\binom{v/2}{n}/\binom{v}{n}]^r < 1$, then $m(n) \leq r$.

We now adjust v so as to give the best upper bound r on $m(n)$. This problem is fairly typical of those encountered when analyzing the asymptotics of the probabilistic method. While our language is informal, the argument may be made rigorous. For small ε we may estimate $1-\varepsilon$ by $e^{-\varepsilon}$. With $\binom{v/2}{n}/\binom{v}{n}$ small we may take

$$r \sim v(\ln 2)\bigg/\left[2\binom{v/2}{n}\bigg/\binom{v}{n}\right] .$$

When v is large, $2\binom{v/2}{n}/\binom{v}{n}$ is roughly 2^{1-n}. (Given $v/2$ Red and $v/2$ Blue balls, $2\binom{v/2}{n}/\binom{v}{n}$ is the probability that a random n-set is monochromatic, whereas 2^{1-n} is the probability that n points independently selected would have the same color.

In probability language, the distinction is between sampling without replacement (the n-set) and sampling with replacement (n points), and for large v the distinction is negligible.) This approximation would yield $r \sim v2^{n-1}(\ln 2)$. As v gets smaller, however, the approximation becomes less accurate and, as we wish to minimize r, the tradeoff becomes essential. We use a second-order approximation

$$2\binom{v/2}{n}\bigg/\binom{v}{n} = 2^{1-n}\prod_{i=0}^{n-1}(v-2i)/(v-i)$$
$$\sim 2^{1-n}\,\mathrm{e}^{-n^2/2v}\,,$$

(estimating $(v-2i)/(v-i) \sim 1-i/v \sim \mathrm{e}^{-i/v}$) so that $r \sim 2^{n-1}(\ln 2)\,\mathrm{e}^{n^2/2v}$. Elementary calculus gives $v = n^2/2$ for the optimal value. We combine both bounds with the expression

$$2^{n-1} \leqslant m(n) < (1+\mathrm{o}(1))[\mathrm{e}(\ln 2)/4]n^2 z^n\,.$$

Written in the above form the asymptotic behavior of $m(n)$ appears essentially solved. This is unfortunate for, as we shall soon see, when viewed from a probabilistic standpoint the gap is indeed wide.

Our second basic weapon is called *linearity of expectation*. It is, quite simply, that if $X_1, \ldots, X_n$ are random variables, then

$$\mathrm{E}[X_1 + \cdots + X_n] = \mathrm{E}[X_1] + \cdots + \mathrm{E}[X_n]\,.$$

The strength of this property lies in the fact that it remains valid even if the X_i are dependent on each other. Consider the *hat-check girl problem*: if n hats are distributed at random to n men, how many get their own hat? Let X_i be the indicator random variable for the event that the ith man gets his own hat back (i.e., $X_i = 1$ if he does, 0 if he does not) and set $X = X_1 + \cdots + X_n$ so that X is the number of men getting their own hat back. The expectation of an indicator random variable is simply the probability of the associated event, so $\mathrm{E}[X_i] = 1/n$. By linearity of expectation, $\mathrm{E}[X] = 1$. On the average, one man will get his own hat back. Of course, this argument tells us nothing about the distribution of X (which in this case is roughly Poisson with mean 1) but often the expectation alone is sufficient for our needs.

Given a round-robin tournament (as defined earlier) on n players, a *Hamiltonian path* is defined as a permutation $P(1), \ldots, P(n)$ of the players so that for $1 \leqslant i \leqslant n-1$ $P(i)$ beats $P(i+1)$. Szele (1943), in arguably the first use of probabilistic methods, asked for the maximal number of Hamiltonian paths a tournament could have. Let X be the number of Hamiltonian paths in a random tournament. Let X_P be the indicator random variable of the event that permutation P generates a Hamiltonian path. Then $X = \sum X_P$, the summation over all $n!$ permutations P. For any given P, $\mathrm{E}[X_P] = 2^{-(n-1)}$, since that is the probability that the $n-1$ games $P(i)$ versus $P(i+1)$ all have the desired outcome. By linearity of expectation, $\mathrm{E}[X] = n!2^{-(n-1)}$. A random variable cannot be always less than its expectation. There must be some point in the probability space with $x \geq n!2^{-(n-1)}$. That is, there is a tournament with at least $n!2^{-(n-1)}$ Hamiltonian paths.

Linearity of expectation is often used for indicator random variables of events of equal probability. Let n events each of probability p be given and let X be the number of these events that occur. Then $\mathrm{E}[X] = np$. Thus there is some point in the probability space for which at most (and, similarly, at least) np of the events occur. When $np < 1$ we may consider the counting sieve as a corollary. We emphasize again that the n events need not be independent for these deductions.

Let us return to 2-colorability. Fix a family $\mathcal{F} = \{S_1, \ldots, S_r\}$ of sets, each of cardinality n, and color the underlying points randomly. Let X_i be the indicator random variable for S_i being monochromatic. Let $X = X_1 + \cdots + X_r$, i.e., the number of monochromatic sets. By linearity of expectation $\mathrm{E}[X] = r2^{1-n}$. When $r < 2^{n-1}$, reformulating Erdős' argument, $\mathrm{E}[X] < 1$ and hence $X = 0$ is not a null event. The inverse does not hold. When $r \geq 2^{n-1}$, $\mathrm{E}[X] \geq 1$, but this, by itself, still allows the possibility that $X = 0$ sometimes – if, say, $\mathrm{E}[X] = 100$ and X had "a lot of distribution" one would naturally expect $X = 0$ to occur with positive probability. We may rephrase determination of $m(n)$ as finding the maximal w so that, in this particular instance, $\mathrm{E}[X] \leq w$ implies $\Pr[X = 0] > 0$. Then $1 \leq w \leq cn^2$, the gap is indeed wide. In section 2 we give a more advanced technique due to Beck (1978) which will show $w > cn^{1/3}$.

The following result plays a central role in extremal set theory (see chapter 24).

Theorem 1.3 (Yamamoto's inequality). *Suppose $\mathcal{F} \subseteq 2^{[n]}$ is an antichain. Then*

$$\sum_{F \in \mathcal{F}} 1 \Big/ \binom{n}{|F|} \leq 1 .$$

Proof. Select a permutation σ uniformly from S_n and let $\mathcal{C}_\sigma$ be the chain it generates. That is,

$$\mathcal{C}_\sigma = \{\{\sigma(j) \colon 1 \leq j \leq i\} \colon 0 \leq i \leq n\} .$$

Let X_F be the indicator random variable for the event $F \in \mathcal{F}$ and set

$$X = \sum_{F \in \mathcal{F}} X_F .$$

The chain $\mathcal{C}_\sigma$ contains exactly one set of size $|F|$, uniformly distributed among all such sets. Hence

$$\mathrm{E}[X_F] = \Pr[F \in \mathcal{C}_\sigma] = 1 \Big/ \binom{n}{|F|}$$

and by linearity of expectation

$$\mathrm{E}[X] = \sum_{F \in \mathcal{F}} 1 \Big/ \binom{n}{|F|} .$$

If Yamamoto's inequality fails, then $\mathrm{E}[X] > 1$ so that there exists a specific σ for which $X \geq 2$, i.e. $|\mathscr{C}_\sigma \cap \mathscr{F}| \geq 2$. But this would imply that $\mathscr{F}$ is not an antichain. □

Certainty Yamamoto's inequality may be proven without reference to probabilistic ideas. Indeed, probabilistic results may always be phrased as counting arguments. Oftentimes (including, arguably, the above example) the probabilistic framework provides key insight into the core of the result.

Our third basic weapon is the bounding of large deviations. We require results of a purely probabilistic nature which, for convenience, are gathered in the Appendix, which is self-contained.

Let $\mathscr{F}$ be a family of sets with underlying point set Ω. A two-coloring will refer to a map $\chi : \Omega \rightarrow \{-1, +1\}$. For any $S \in \mathscr{F}$ we set

$$\chi(S) = \sum_{i \in S} \chi(i) ,$$

so that, for example, $\chi(S) = 0$ means S is colored equally $(+1)$ and (-1). The *discrepancy* of χ is defined by

$$\operatorname{disc}(\chi, \mathscr{F}) = \max_{S \in \mathscr{F}} |\chi(S)|$$

and the discrepancy of $\mathscr{F}$, denoted by $\Delta(\mathscr{F})$, is given by

$$\Delta(\mathscr{F}) = \min \operatorname{disc}(\chi, \mathscr{F}) ,$$

the minimum over all two-colorings χ. Thus $\Delta(\mathscr{F}) \leq K$ means that it is possible to two-color Ω so that all $S \in \mathscr{F}$ are "balanced" within K. See chapter 27 for a general discussion of discrepancy.

Now suppose $\mathscr{F}$ consists of n sets on an underlying set Ω of n points. We use large deviations to bound $\Delta(\mathscr{F})$. Consider a random two-coloring χ and for each $S \in \mathscr{F}$ let A_S be the event that $|\chi(S)| > \alpha$. When S has m elements, $\chi(S)$ has distribution S_m, so, by Corollary A.2,

$$\Pr[A_S] < 2\,\mathrm{e}^{-\alpha^2/2m} \leq 2\,\mathrm{e}^{-\alpha^2/2n} .$$

When $2n\,\mathrm{e}^{-\alpha^2/2n} < 1$, the counting sieve implies $\wedge \bar{A}_S \neq \emptyset$, there is a χ so that all $|\chi(S)| \leq \alpha$ and so $\Delta(\mathscr{F}) \leq \alpha$. Solving for α,

$$\Delta(\mathscr{F}) \leq \sqrt{2n \ln(2n)} .$$

This result may be "reversed", analogously to the $m(n)$ bounds, to show the existence of an $\mathscr{F}$ with n sets on n elements with $\Delta(\mathscr{F}) > c\sqrt{n}$, c an absolute constant. An improvement on the upper bound is given in section 2.

For the next example, we return again to tournaments. Given a (round-robin) tournament T on n players and an ordering $P(1), \ldots, P(n)$ of the players, the *fit* $f(T, P)$ is the number of pairs of players i, j, with $i < j$ and $P(i) < P(j)$. That is, the fit is the number of games for which the ordering is "correct". Set $m = \binom{n}{2}$ for convenience. If T is transitive (has a natural order), then $f(T, P) = m$ for that ordering. Let $g(T)$ equal the maximum of $f(T, P)$ over all P, i.e., the best fit. For any T we may try an arbitrary P and its reverse. One of them will agree with T in at least half the games. Hence $g(T) \geqslant m/2$. Erdős and Moon (1965) showed that this is nearly best possible. Consider a random tournament T. For each P, $f(T, P)$ has binomial distribution $\mathrm{B}(m, \frac{1}{2})$ since each game has independent probability $\frac{1}{2}$ to agree with the ordering P. Thus $2f(T, P) - m$ has distribution S_m. Let A_P be the event that $f(T, P) \geqslant m/2 + a$. Then, by Corollary A.2,

$$\Pr[A_P] = \Pr[S_m \geqslant 2a] < \mathrm{e}^{-(2a)^2/2m} .$$

We desire $\wedge \bar{A}_P \neq \emptyset$, but we have an "enormous" number, $n!$, of P. Fortunately, the tail of the distribution S_m decreases very quickly. We set $a = n^{3/2}(\ln n)^{1/2}/2$ so that $\Pr[A_P] < n^{-n} < (n!)^{-1}$. (This use of what normally would be considered the extreme tail of the distribution is quite common.) There exists T on n players with

$$g(T) < \frac{1}{2}\binom{n}{2} + n^{3/2}(\ln n)^{1/2}/2 .$$

For the mathematician, even combinatorialist, not familiar with probabilistic methods a much weaker method can be quite impressive: there exist tournaments which cannot be ranked so that more than 51% of the games are in order. Construction of specific tournaments with this property appears to be quite difficult and quite possibly, in this author's opinion, impossible. (This leads, however, into a logical thicket around the terms "construction" and "specific" in which we shall not become embroiled!) Let us set $h(n)$ equal to the minimum of $g(T)$ over all tournaments T on n players so that we have shown

$$\frac{1}{2}\binom{n}{2} \leqslant h(n) \leqslant \frac{1}{2}\binom{n}{2} + \frac{1}{2}n^{3/2}(\ln n)^{1/2} .$$

From the probabilistic standpoint the $\frac{1}{2}\binom{n}{2}$ term may be regarded as the "zero term" and so this inequality could certainly stand improvement, as indeed it will get in section 4.

2. Advanced methods

In section 1 we created probability spaces and showed the existence of a good point (configuration, coloring, tournament) by showing that the bad points have measure less than one. By moving slightly away from the threshold value we find that the bad points are negligible in measure so that a randomly selected point will almost surely be good. We consider a method "advanced" if it enables us to find "rare" points, i.e., if it works even when the set of good points is very small in measure. An advanced method allows us to find (or at least to prove the existence of) a needle in a haystack, an elementary method shows us the hay. As with customary mathematical usage, advanced methods can be quite simple and elementary methods most ingenious. For the latter consider de la Vega's improvement, given in section 4, of the ranking of tournaments function $h(n)$. For the former . . . read on.

The deletion method takes a random configuration which is not good but is bad in only a few places. An alteration, generally a deletion, is made at each bad place until the object, now generally smaller, has the desired property. Consider a random two-coloring of the edges of K_n and let X denote the number of monochromatic k-sets. Then $\mathrm{E}[X] = \binom{n}{k}2^{1-\binom{k}{2}}$ by linearity of expectation. To bound the Ramsey function $R(k, k)$ by elementary methods we took n such that $\mathrm{E}[X] < 1$. Now take n larger so that $\mathrm{E}[X] = w > 1$. There is a coloring with at most w monochromatic k-sets. Select one vertex arbitrarily from each such k-set and delete it from the vertex set. What remains is at least $n - w$ vertices with no monochromatic k-set and hence $R(k, k) \geq n - w$. The value $n \sim (1/\mathrm{e}\sqrt{2})k2^{k/2}$ gave $w \sim 1$. We may increase n by a factor of nearly $\sqrt{2}$ and still have $w \ll n$. The monochromatic k-sets are then small in number and their deletion still leaves roughly n vertices. Thus

$$R(k, k) > \frac{1}{\mathrm{e}} k2^{k/2}(1 + \mathrm{o}(1)) .$$

Let G be a graph with n vertices and $nt/2$ edges, $t \geq 1$. Let $\alpha(G)$ denote the maximal size of an independent set. Turán's Theorem (see chapter 22) gives $\alpha(G) \geq n/(t + 1)$. With the deletion method we shall get a set half that size. Fix the graph G. (Note that G itself is not random.) Let p be, for the moment, arbitrary and define a random set S of vertices by placing each vertex of G in S with independent probability p. Let X be the size of S and let Y be the number of edges of G with both vertices in S. Clearly $\mathrm{E}[X] = np$. Since each of the $nt/2$ edges of G has probability p^2 of having both vertices in S, linearity of expectation gives $\mathrm{E}[Y] = (nt/2)p^2$ and hence

$$\mathrm{E}[X - Y] = np - (nt/2)p^2 .$$

Now set $p = 1/t$ to maximize this quantity: $\mathrm{E}[X - Y] = n/2t$. There is a "point" in the probability space – a specific set S – for which $X - Y \geq n/2t$. Delete one vertex arbitrarily from each edge of S. This leaves a set S^* which is independent and has at least $X - Y$ vertices. Thus $\alpha(G) \geq n/2t$.

Let X be a random variable with mean μ and variance σ^2. The basic inequality of Chebyshev states that for any positive λ

$$\Pr[|X - \mu| > \lambda\sigma] < 1/\lambda^2 .$$

Taking $\lambda = \varepsilon\mu/\sigma$,

$$\Pr[|X - \mu| > \varepsilon\mu] < \sigma^2/\varepsilon^2\mu^2 .$$

In many examples X depends on a parameter approaching infinity. If we know that $\mathrm{Var}[X] = \mathrm{o}(\mathrm{E}[X]^2)$, then for any fixed ε the above probability is $\mathrm{o}(1)$. This is the *second-moment method*: if $\mathrm{Var}[X] = \mathrm{o}(\mathrm{E}[X]^2)$, then X is almost always almost equal to its expectation. (This method is particularly valuable in studying random graphs, see chapter 6.)

As an example, let G be a regular graph of degree t on n vertices with $1 \ll t \ll n$. Set $p = 1/t$ and, as before, define a random set S of vertices by placing each vertex of G in S with independent probability p. Again, let X be the size of S and Y be the number of edges of G with both vertices in S. X has mean np and variance $np(1-p)$ so by the second-moment method $X \sim n/t$ almost always. (However, in this case the bounds of the Appendix – or standard bounds on the binomial distribution – give much stronger results.) Let $1 \leqslant i \leqslant m = nt/2$ index the edges of G and let Y_i be the indicator random variable for the ith edge having both vertices in S. Then $\mathrm{E}[Y_i] = p^2$ and $Y = Y_1 + \cdots + Y_m$. We employ the general formula

$$\mathrm{Var}[Y] = \sum_{i,j=1}^{m} \mathrm{Cov}[Y_i, Y_j]\,,$$

where $\mathrm{Cov}[Y_i, Y_j] = \mathrm{E}[Y_iY_j] - \mathrm{E}[Y_i]\mathrm{E}[Y_j]$ is the covariance. Critically, when two random variables are independent their covariance is zero. This is the case when the ith and jth edges have no vertex in common. Each edge has a vertex in common with $2t-1$ edges (including itself). In those cases we use the weak bound

$$\mathrm{Var}[Y_i, Y_j] \leqslant \mathrm{E}[Y_iY_j] \leqslant \mathrm{E}[Y_i] = p^2\,.$$

Thus $\mathrm{Var}[Y] \leqslant m(2t-1)p^2$. But $\mathrm{E}[Y]^2 = m^2p^2 \gg m(2t-1)p^2$ so the second-moment method applies and $Y \sim \mathrm{E}[Y]$ almost always. Combining these results, S almost always has asymptotically n/t vertices and $n/2t$ edges. Therefore, there will exist specific S with asymptotically n/t vertices *and* $n/2t$ edges.

Let $t < k < n$. Let $\mathscr{F}$ be a family of k-sets with underlying point set $[n]$. $\mathscr{F}$ is called a $t-(n,k,1)$ design if every t-set T is contained in a unique $K \in \mathscr{F}$. The existence of designs is a major combinatorial question (see chapter 14) to which probabilistic methods have not yet been able to contribute. Let $M(n,k,t)$ denote the minimal cardinality of a family $\mathscr{F}$ that *covers* all t-sets, i.e., that every t-set is contained in at least one $K \in \mathscr{F}$. A simple counting argument shows $M(n,k,t) \geqslant \binom{n}{t}/\binom{k}{t}$ with equality if and only if a $t-(n,k,1)$ design exists. Let t,k be fixed and n approach infinity. We find asymptotic upper bounds on $M(n,k,t)$.

Let $\mathscr{F}$ be a random collection where each k-set K is placed in $\mathscr{F}$ with independent probability p. It is convenient to parametrize $p = [x\binom{n}{t}/\binom{k}{t}]/\binom{n}{k} = x/\binom{n-t}{k-t}$ so that $\mathrm{E}[|\mathscr{F}|] = p\binom{n}{k} = x\binom{n}{t}/\binom{k}{t}$. For each t-set T let A_T denote the event that $\mathscr{F}$ does not cover T. As T is contained in $\binom{n-t}{k-t}$ k-sets,

$$\Pr[A_T] = (1-p)^{\binom{n-t}{k-t}} \sim \mathrm{e}^{-x}\,.$$

To use elementary methods we set $x = (1+\varepsilon)\ln[\binom{n}{t}]$, so that $\Pr[A_T] \ll 1/\binom{n}{t}$. Then $\wedge \bar{A}_T$ almost always, i.e., $\mathscr{F}$ almost always covers all t-sets. By the second-moment method $|\mathscr{F}| \sim \binom{n}{k}p$ almost always. Therefore, there will be a specific $\mathcal{F}$ with $|\mathcal{F}| \sim \binom{n}{k}p$ which covers all t-sets. As this holds for any $\varepsilon > 0$,

$$M(n, k, t) < (1 + o(1))\left[\binom{n}{t}\Big/\binom{k}{t}\right] \ln\left[\binom{n}{t}\right].$$

The deletion method (though here it involves enlarging a configuration) allows a substantial improvement. With random $\mathcal{F}$ as previously defined, let Z denote the number of t-sets T not covered. As $\Pr[A_T] \sim e^{-x}$, linearity of expectation gives $E[Z] \sim \binom{n}{t} e^{-x}$. For any positive random variable Z and any positive δ

$$\Pr[Z \geq (1 + \delta)E[Z]] < 1/(1 + \delta).$$

This general principle can be quite useful. Roughly, Z cannot almost always be asymptotically greater than its expectation. We know $|\mathcal{F}| \sim \binom{n}{k}p = x\binom{n}{t}/\binom{k}{t}$ almost always. Thus with positive probability $|\mathcal{F}| \sim x\binom{n}{t}/\binom{k}{t}$ *and* fewer than $(1 + \delta)\binom{n}{t} e^{-x}$ t-sets are not covered by $\mathcal{F}$. Fix such an $\mathcal{F}$. For each t-set T not covered add to $\mathcal{F}$ an arbitrary k-set K containing T. The new family $\mathcal{F}^*$ now covers all the t-sets (we have "corrected the errors") and has size at most roughly $\binom{n}{t}[x/\binom{k}{t} + e^{-x}]$. Elementary calculus gives the value $x = \ln[\binom{k}{t}]$ to optimize this quantity. Thus

$$M(n, k, t) < \left[\binom{n}{t}\binom{k}{t}\right]\left(1 + \ln\left[\binom{k}{t}\right] + o(1)\right).$$

The tale is not over. Erdős and Hanani (1963) conjectured that for k, t fixed

$$\lim M(n, k, t)\binom{k}{t}\Big/\binom{n}{t} = 1.$$

This would mean that asymptotically one could find families $\mathcal{F}$ which were "nearly" designs. It was twenty years before Rödl (1985) found a proof of this conjecture. The proof technique, called the *Rödl nibble*, is most ingenious. We describe it in a more general context. Let k be fixed and let G be a k-graph on n vertices, each vertex having degree asymptotically D. Here, $D = D(n)$ approaches infinity. Suppose further that every two vertices have only $o(D)$ edges in common. Then Frankl and Rödl (1985) show that there is a set of $\sim n/k$ edges covering all the vertices.

Let $\varepsilon > 0$ be fixed. Choose edges independently with probability $p = \varepsilon/D$ so that $\sim n\varepsilon/k$ edges are selected. Each vertex lies in $\sim D$ edges and so has probability $\sim (1 - p)^D \sim e^{-\varepsilon}$ of not being covered. A second-moment argument, using that two vertices have $o(D)$ edges in common, gives that $\sim n\, e^{-\varepsilon}$ vertices are not covered almost always. Let G' be the restriction of G to the uncovered vertices. One can also show (and this is the hardest part) that G' has appropriate asymptotic regularity. This allows the nibble to be iterated and the iteration is continued until fewer than εn vertices are uncovered. These are covered one by one as with the deletion method.

The Rödl nibble uses $\sim n\varepsilon/k$ edges to cover $\sim n(1 - e^{-\varepsilon})$ vertices. For ε very small $1 - e^{-\varepsilon} \sim \varepsilon$, so the covering is very efficient. The final one-by-one covering is highly inefficient (we want each edge to cover k new vertices and here it covers only one) but involves only εn edges. The total number of edges used may be written in the form $(n/t)f(\varepsilon)$. While $f(\varepsilon) > 1$ for any fixed ε, $\lim_{\varepsilon \to 0} f(\varepsilon) = 1$ and hence $\sim n/t$ edges suffice to cover all the vertices.

The *recoloration method* is similar to the deletion method. One begins with a random coloring which yields a few bad spots. At these spots a random recoloration is made. The recoloration must be strong enough to destroy all bad spots yet not so strong as to create new bad spots. We give an improvement due to Beck (1978) of the lower bound $m(n) \geqslant 2^{n-1}$ discussed in section 1.

Let F be a collection on n-sets of cardinality $2^{n-1}t$. A random coloring yields an expected number t of monochromatic sets. We recolor by taking every point x lying in at least one monochromatic set and switching its color with probability p. These switches are independent over the x. Note that it is immaterial whether x lies in one or fifty monochromatic sets. Given that S was, say, Red it remains Red with probability $(1-p)^n \sim \mathrm{e}^{-pn}$. We set $p = (1+\varepsilon)(\ln t)/n$ so that the expected number of such S is roughly $(2^{n-1}t)(2^{1-n})\,\mathrm{e}^{-pn} = t^{-\varepsilon} \ll 1$. Almost always the recoloration fixes all bad spots. Does it create any new ones? One case is simple: almost always no set is monochromatic in the first coloring and is switched to the other color in the recoloring.

What about S that were not monochromatic but became, say, Red in the recoloring. For $S, T \in F$ with $S \cap T \neq \emptyset$ and $V \subseteq S - T$, $V \neq S - T$, let A_{STV} be the event that T was Blue in the first coloring, S was Red except for precisely $(S \cap T) \cup V$, and S became completely Red in the recoloring. If S became Red then some A_{STV} must have occurred. We restrict our attention to the case $|S \cap T| = 1$; the other cases are similar and give lower probabilities. Let V have cardinality v.

For event A_{STV} to occur, the original coloring of $S \cup T$ must be precisely correct and a given $v+1$ points must have their colors switched in the recoloring. Thus $\Pr[A_{STV}] \leqslant 2^{-2n+1}p^{v+1}$. (This may be a gross overestimate as in addition all points of V must be switchable – must lie in sets monochromatic in the first coloring. It appears difficult to utilize this condition.) There are $\binom{n-1}{v} < n^v/v!$ choices of V with v elements. As v increases the larger number of choices for V is balanced by the smaller probability that all points of V are switched in the recoloring. Let A_{ST} be the disjunction of A_{STV} over all V. Then

$$\begin{aligned}\Pr[A_{ST}] &\leqslant 2^{-2n+1} \sum_{v=0}^{n-2} (n^v/v!)p^{v+1} \\ &\leqslant 2^{-2n+1} p \sum_{v=0}^{\infty} (np)^v/v! \\ &= 2^{-2n+1} p\, \mathrm{e}^{np} = 2^{-2n} t^{1+o(1)}/n\,.\end{aligned}$$

(When $|S \cap T| > 1$ an even smaller bound is found.) There are at most $(2^{n-1}t)^2$ pairs S, T so

$$\Pr\left[\bigvee A_{ST}\right] \leqslant 2^{2n-2}t^2 2^{-2n} t^{1+o(1)}/n = t^{3+o(1)}/n\,.$$

As long as $t < n^{1/3-o(1)}$ this quantity is o(1) and so the recoloring is almost always free of monochromatic sets. Hence $m(n) > n^{1/3-o(1)}2^n$. With more attention to detail $m(n) > cn^{1/3}2^n$ has been shown.

The counting sieve is in one sense best possible. When events A_i are pairwise disjoint the condition $\sum \Pr[A_i] < 1$ cannot be weakened. At the opposite extreme, if the events A_i are mutually independent, then we only need that all $\Pr[A_i] < 1$ to assure $\wedge \bar{A}_i \neq \emptyset$. The *Lovász sieve* is employed when there is "much independence" among the A_i. Let $A_1, \ldots, A_n$ be events. A graph G on the index set $\{1, \ldots, n\}$ is said to be a dependency graph if, for all i, A_i is mutually independent of $\{A_j: \{i, j\} \not\in E(G)\}$. We emphasize that A_i must not only be independent of each such A_j individually but also must be independent of any Boolean combination of the A_j. The dependency graph is not uniquely determined by the events. In application, however, there will usually be a natural choice of a dependency graph.

Lovász Local Lemma 2.1 (Erdős and Lovász 1975). *Let G be a dependency graph on events $A_1, \ldots, A_n$. Assume each point of G has degree at most d. Assume $\Pr[A_i] \leqslant p$ for each i. Assume $4dp \leqslant 1$. Then $\wedge \bar{A}_i \neq \emptyset$.*

Proof. We prove by induction on m that for any m events (calling them $A_1, \ldots, A_m$ only for convenience)

$$\Pr[A_1 \mid \bar{A}_2 \cdots \bar{A}_m] \leqslant 1/2d \,.$$

For $m = 1$ this is obvious. Let $\{2, \ldots, q\}$ be the points of $\{2, \ldots, m\}$ adjacent to 1 in the dependency graph G. Then

$$\Pr[A_1 \mid \bar{A}_2 \cdots \bar{A}_m] = \frac{\Pr[A_1 \bar{A}_2 \cdots \bar{A}_q \mid \bar{A}_{q+1} \cdots \bar{A}_m]}{\Pr[\bar{A}_2 \cdots \bar{A}_q \mid \bar{A}_{q+1} \cdots \bar{A}_m]} \,.$$

We bound the numerator

$$\begin{aligned}\Pr[A_1 \bar{A}_2 \cdots \bar{A}_q \mid \bar{A}_{q+1} \cdots \bar{A}_m] &\leqslant \Pr[A_1 \mid \bar{A}_{q+1} \cdots \bar{A}_m] \\ &= \Pr[A_1] \leqslant 1/4d \,,\end{aligned}$$

since A_1 is mutually independent of $\{A_{q+1}, \ldots, A_m\}$. The denominator is bounded by

$$\begin{aligned}\Pr[\bar{A}_2 \cdots \bar{A}_q \mid \bar{A}_{q+1} \cdots \bar{A}_m] &\geqslant 1 - \sum_{i=2}^{q} \Pr[A_i \mid \bar{A}_{q+1} \cdots \bar{A}_m] \\ &\geqslant 1 - (q-1)(1/2d) \quad \text{(induction)} \,,\end{aligned}$$

which is at least $\frac{1}{2}$ since $q - 1 \leqslant d$. Thus

$$\Pr[A_1 \mid \bar{A}_2 \cdots \bar{A}_m] \leqslant (1/4d)/(1/2) = 1/2d \,,$$

completing the induction. Finally,

$$\Pr[\bar{A}_1 \cdots \bar{A}_n] = \prod_{i=1}^{n} \Pr[\bar{A}_i \mid \bar{A}_1 \cdots \bar{A}_{i-1}] \geqslant (1 - 1/2d)^n > 0 . \qquad \square$$

A striking feature of the Lovász sieve is the lack of condition on the total number n of events. When n is large, $\Pr[\wedge \bar{A}_i]$ can be very small and the Lovász local Lemma sieves out a needle from the haystack.

Consider the lower bound to the Ramsey function $R(k, k)$. We two-color K_n randomly. For each set S of k vertices let A_S be the event that S is monochromatic. We define the discrepancy graph G by making S, T adjacent if and only if S and T have an edge in common, i.e., $|S \cap T| \geqslant 2$. The joint veracity of the events A_T with T not adjacent to S cannot affect $\Pr[A_S]$ since S has different edges and all edges are colored independently. We bound d, the number of T with $|S \cap T| \geqslant 2$, by $d \leqslant \binom{k}{2}\binom{n}{k-2}$. The Lovász sieve gives that if $4\binom{k}{2}\binom{n}{k-2} \leqslant 1$, then $R(k, k) > n$. Some calculation shows

$$R(k, k) \geqslant \frac{\sqrt{2}}{e} k 2^{k/2}(1 + o(1)) .$$

This improves elementary methods by a factor of two. It is the best lower bound for $R(k, k)$ currently known. With the upper bound roughly 4^k, the improvement, sadly, does not really help in finding the true order of this basic combinatorial function.

The Lovász sieve is most striking when the degree d is fixed and the number n of events can be arbitrarily large.

Theorem 2.2. *Let k, m be positive integers with $4km(m - 1)(1 - 1/k)^m < 1$. Let S be any subset of $\mathbb{R}$, the real numbers, with m elements. Then there exists a k-coloring $\chi : \mathbb{R} \rightarrow \{1, \ldots, k\}$ such that for any $t \in \mathbb{R}$ the translate $S + t$ is colored with all k colors.*

Gallai's Theorem, a generalization of van der Waerden's Theorem, gives that for any finite $S \subseteq \mathbb{R}$ and any k-coloring $\chi : \mathbb{R} \rightarrow \{1, \ldots, k\}$ there exists S' homothetic to S (i.e., $S' = aS + t$) which is monochromatic. When homothety is replaced by translation the above theorem gives a powerful result in the opposite direction (see chapter 25).

Proof. We first show that any finite $D \subseteq \mathbb{R}$ may be k-colored so that all translates $S + t$ entirely contained in D have all k colors. Color D randomly, i.e., for each $x \in D$ select $\chi(x)$ by flipping a fair k-faced die. For each t such that $S + t \subseteq D$ let A_t denote the event that $S + t$ does not have all k colors. Clearly, $\Pr[A_t] \leqslant k(1 - 1/k)^m$. We define a dependency graph G, letting t, t' be adjacent if and only if $(S + t) \cap (S + t') \neq \emptyset$, i.e., $t' = t + s' - s''$ for some s', s'' in S. G has degree at most $d = m(m - 1)$. The Lovász Local Lemma gives $\wedge \bar{A}_t \neq \emptyset$, so there is a coloring χ of D as desired.

Extending colorings of arbitrary finite D to a coloring of all of $\mathbb{R}$ requires a new concept: the *compactness principle*. Quite generally, let Ω be an infinite set and suppose $\mathcal{U}$ is a family of pairs (B, χ), $B \subseteq \Omega$, B finite, χ a k-coloring of B, i.e., $\chi : B \rightarrow \{1, \ldots, k\}$. Suppose $\mathcal{U}$ is closed under restriction – if $(B, \chi) \in \mathcal{U}$ and $C \subseteq B$, then $(C, \chi|_C) \in \mathcal{U}$. Suppose also that for all finite $B \subseteq \Omega$ there is a χ with $(B, \chi) \in \mathcal{U}$. The compactness principle is that there then exists a $\chi : \Omega \rightarrow \{1, \ldots, k\}$ such that for all $B \subseteq \Omega$ $(B, \chi|_B) \in \mathcal{U}$ (see chapter 42). In our case let $\mathcal{U}$ be the family of (D, χ) with all $S + t \subseteq D$ having all k colors. The compactness principle gives a k-coloring χ of $\mathbb{R}$. For any $t(S + t, \chi|_{S+t}) \in \mathcal{U}$, so that $S + t$ has all k colors, completing the proof. □

When events A_i are not symmetric a more general form of the Lovász sieve is appropriate.

Lovász Local Lemma 2.3 (general form). *Let $A_1, \ldots, A_n$ be events with a dependency graph G. Suppose there exist $x_1, \ldots, x_n$, $0 < x_i < 1$, so that for all i*

$$\Pr[A_i] < x_i \prod (1 - x_j) \quad [\text{product over } \{i, j\} \in E(G)] .$$

Then $\wedge \bar{A}_i \neq \emptyset$.

Proof. The induction hypothesis of the earlier proof is replaced by $\Pr[A_1 \mid \bar{A}_2 \cdots \bar{A}_m] \leq x_i$, and the denominator $\Pr[\bar{A}_2 \cdots \bar{A}_q \mid \bar{A}_{q+1} \cdots \bar{A}_m]$ is set equal to $\prod_{i=2}^{q} \Pr[\bar{A}_i \mid \bar{A}_{i+1} \cdots \bar{A}_m]$, which is bounded by the induction hypothesis. We omit the full proof. □

Setting $y_i = x_i / \Pr[A_i]$ the condition of Lemma 2.3 becomes

$$\ln y_i > \sum_{\{i,j\} \in G} - \ln(1 - y_j \Pr[A_j]) .$$

When all $y_j \Pr[A_j] \ll 1$ this may be simplified, approximately, to

$$\ln y_i > \sum_{\{i,j\} \in G} y_j \Pr[A_j] .$$

Consider the lower bound on the Ramsey function $R(3, k)$. Color K_n by letting each edge be Red (the first color) with probability p, otherwise Blue. For each 3-set S let A_S be the event that S is Red, for each k-set T let B_T be the event that T is Blue. A dependency graph G is given by joining two indices if they intersect in at least two vertices. There are $\binom{n}{k} \leq (n\,\mathrm{e}/k)^k$ k-sets T. Each 3-set S is adjacent to less than $3n$ other 3-sets [a critical improvement over $\binom{n}{3}$] and each k-set T is adjacent to less than $\frac{1}{2}k^2 n$ 3-sets S. Associate to each A_S the same $y_S = y$ and to each B_T the same $y_T = z$. If, approximately, there exist p, y, z with

$$p < 1\,, \qquad y\,\Pr[A_S] \ll 1\,, \qquad z\,\Pr[B_T] \ll 1\,,$$
$$\ln y > y\,\Pr[A_S](3n) + z\,\Pr[B_T](n\,\mathrm{e}/k)^k\,,$$
$$\ln z > y\,\Pr[A_S]k^2n/2 + z\,\Pr[B_T](n\,\mathrm{e}/k)^k\,,$$

then $R(3, k) > n$. Erdős (1961), in one of the early masterpieces of the probabilistic method, proved $R(3, k) > ck^2/(\ln k)^2$ by a delicate use of the deletion method. With the Lovász sieve this same result falls out with some prosaic calculation – set $p = c_1 n^{-1/2}$, $k = c_2 n^{1/2} \ln n$, $z = \exp[c_3 n^{1/2}(\ln n)^2]$ and $y = 1 + c_4$ for appropriate values of the constants.

Our final method involves the use of *entropy*. Let a random variable Y assume m possible values with probabilities $p_1, \ldots, p_m$. The entropy of Y, denoted by $H[Y]$, is given by

$$H[Y] = \sum_{i=1}^{m} -p_i \log_2 p_i\,.$$

Entropy measures the information content of Y. If $m = 2^t$ and all $p_i = 2^{-t}$, then $H[Y] = t$. Convexity arguments give that if the p_i are smaller the entropy must be larger. In contrapositive form we express this as a *concentration property*: if $H[Y] \leq t$, then there is some b for which $\Pr[Y = b] \geq 2^{-t}$. Let $Y_1, \ldots, Y_s$ be random variables, not necessarily independent, on a common space. Entropy satisfies the *subadditivity property*:

$$H[(Y_1, \ldots, Y_s)] \leq H[Y_1] + \cdots + H[Y_s]\,.$$

Let $\mathscr{F} = \{S_1, \ldots, S_n\}$ be a family of n sets on underlying point set $\Omega = \{1, \ldots, n\}$. We reconsider the problem of bounding the discrepancy $\Delta(\mathscr{F})$. Let $\chi : \Omega \to \{-1, +1\}$ be random. For $1 \leq j \leq n$ let Y_j be the nearest integer to $\chi(S_j)/20n^{1/2}$. $Y_j = 0$ when $\chi(S_j) \leq 10n^{1/2}$, so, by Corollary A.2,

$$\Pr[Y_j = 0] > 1 - 2\,\mathrm{e}^{-50}\,.$$

Let $s > 0$. $Y_j = s$ requires $\chi(S_j) > (s - \frac{1}{2})(20n^{1/2})$, so by Theorem A.1,

$$\Pr[Y_j = s] < \mathrm{e}^{-50(2s-1)^2}\,.$$

$\Pr[Y_j = -s]$ has the same bound. We bound the entropy by the infinite sum

$$H[Y_j] \leq -(1 - 2\,\mathrm{e}^{-50})\log_2(1 - 2\,\mathrm{e}^{-50}) - 2(\mathrm{e}^{-50}\log_2 \mathrm{e}^{-50} + \cdots) = c\,,$$

where calculation gives $c < 3 \times 10^{-20}$. The subadditivity property gives

$$H[(Y_1, \ldots, Y_n)] \leq cn$$

and now the concentration property gives

$$\Pr[(Y_1, \ldots, Y_n) = (b_1, \ldots, b_n)] \geq 2^{-cn} .$$

Fix these b's and let $\mathscr{C}$ be the set of χ giving those b's. As the probability distribution for χ was uniform over the 2^n possible colorings, $|C| \geq 2^{n(1-c)}$.

We can associate colorings $\chi : \Omega \rightarrow \{-1, +1\}$ with points $(\chi(1), \ldots, \chi(n))$ on the Hamming n-cube $C^n = \{-1, +1\}^n$. The Hamming distance (see chapter 16) between $\chi_1, \chi_2 \in C^n$ is then the number of x for which $\chi_1(x) \neq \chi_2(x)$. A celebrated result of Kleitman (1966) says that if $C \subseteq C^n$ has fixed size, then the diameter of C is minimized when C is a ball. Some calculation gives that $\mathscr{C}$ must therefore have diameter at least $n(1 - 10^{-10})$. Pick $\chi_1, \chi_2 \in \mathscr{C}$ at this maximal distance.

Now set $\chi = (\chi_1 - \chi_2)/2$. The possible values for $\chi(x)$ are $+1$, -1, and 0. We think of χ as a partial coloring – when $\chi(x) = 0$ x is uncolored. This occurs exactly when $\chi_1(x) = \chi_2(x)$, which holds for at most $10^{-10}n$ vertices x. For each j the values $\chi_1(S_j)$ and $\chi_2(S_j)$ lie on a common interval of length $20n^{1/2}$ since, critically, their values Y_j are the same. Thus

$$|\chi(S_j)| = |\chi_1(S_j) - \chi_2(S_j)|/2 \leq 10\sqrt{n} .$$

That is, there is a partial coloring of all but at most $10^{-10}n$ vertices with discrepancy at most $10n^{1/2}$. By iterating this process, and tightening the calculation I have shown the following result.

Theorem 2.4 (Spencer 1985). *Let $\mathscr{F}$ be a family of n sets on n points. Then $\Delta(\mathscr{F}) < 6\sqrt{n}$.*

Further results are discussed in chapter 26.

3. Two applications in computer science

We give here two ingenious probabilistic arguments and discuss, briefly, their application to computer science. In both cases our chronology is backwards: it was the computer science problem that motivated the probabilistic approach. Karp (1976) discusses probabilistic analysis of algorithms.

Theorem 3.1. *Let $|V| = n$ and let $\mathscr{H}$ be an arbitrary nonempty family of subsets of V. Let $w : V \rightarrow \{1, \ldots, 2n\}$ be a random function, each $w(v)$ independently chosen uniformly over the range. For $E \in \mathscr{H}$ set $w(E) = \sum_{a \in E} w(a)$. When $\min_{E \in \mathscr{H}} w(E)$ is achieved by a unique $E \in \mathscr{H}$ we say w has a* unique minimum. *Then*

$$\Pr[w \text{ has a unique minimum}] > \tfrac{1}{2} .$$

Remark. When $\mathscr{H}$ is exponential the *pigeonhole principle* insures that there will be many E with equal $w(E)$.

Proof. For $x \in V$ set

$$\alpha(x) = -\min_{x\in E;\, E\in\mathscr{H}} w(E-x) + \min_{x\notin E;\, E\in\mathscr{H}} w(E)\,.$$

Observe that evaluation of $\alpha(x)$ does not require knowledge of $w(x)$. As $w(x)$ is selected uniformly,

$$\Pr[\alpha(x) = w(x)] \leq 1/2n\,,$$

so that

$$\Pr[\alpha(x) = w(x) \text{ for some } x \in V] \leq n/2n = \tfrac{1}{2}\,.$$

But if w had two minimal sets $E_1, E_2 \in \mathscr{H}$ and $x \in E_1 - E_2$, then

$$\min_{x\notin E;\, E\in\mathscr{H}} w(E) = w(E_2)\,,$$

$$\min_{x\in E;\, E\in\mathscr{H}} w(E-x) = w(E_1) - w(x)\,,$$

so that $\alpha(x) = w(x)$. □

Mulmuley et al. (1987) prove this result and use it to give an RNC^2 algorithm to construct a perfect matching in a graph. (See chapter 29 for further discussion.) Let us consider the simpler case when G is a bipartite graph given by an $n \times n$ incidence matrix $A = [a_{ij}]$. Set $V = \{(i,j)\colon a_{ij} = 1\}$, the edges of the graph. Let $w: V \rightarrow \{1, \ldots, 2|V|\}$ be random and define an $n \times n$ matrix $R = [r_{ij}]$ by $r_{ij} = 2^{w(i,j)}$ if $(i,j) \in V$; 0, otherwise. Apply known NC^2 algorithms to calculate $\det(R)$ and the cofactor matrix $C = [c_{ij}]$. Let W be maximal so that $2^W \mid \det(R)$. Let $\mathscr{H}$ denote the set of all perfect matchings $E_\sigma = \{(i, \sigma i)\colon 1 \leq i \leq n\}$, where σ is a permutation on $[n]$ with all $(i, \sigma i) \in V$. Then

$$\det(R) = \sum_{E_\sigma \in \mathscr{H}} \operatorname{sgn}(\sigma) 2^{w(E_\sigma)}\,.$$

Let M be the set of $(i,j) \in V$ so that 2^W is the maximal power of 2 dividing $r_{ij}c_{ij}$. With probability at least $\frac{1}{2}$, $w(E_\sigma)$ has a unique minimum $w(E)$. In that case

$$\det(R) \equiv 2^{w(E)} \bmod 2^{w(E)+1}\,,$$

so $W = w(E)$. Also $r_{ij}c_{ij}$ has a $\pm 2^W$ addend if and only if $(i,j) \in E$, so that $M = E$. Check if M is a perfect matching and output it if so.

We now turn to a result of Yao (1985) on the complexity of the parity function. We adapt the approach of Hastad (1988) which greatly simplifies the proof and makes explicit the probabilistic underpinnings.

Consider Boolean functions G on n variables $x_1, \ldots, x_n$, where 0, 1, $\wedge$, $\vee$, and $\bar{A}$ have their usual meanings of FALSE, TRUE, AND, OR, and NOT A, respectively. The functions x_i, $\bar{x}_i$ are called *atoms*. A *restriction* is a map $\rho: \{1, \ldots, n\} \rightarrow \{0, 1, *\}$; the restriction of G by ρ, denoted $G|_\rho$, is the Boolean function given by setting x_i equal to 0 or 1 or leaving it a variable depending on

$\rho(i)$. For example, with $\rho(1)=0$, $\rho(2)=1$, $\rho(3)=*$, and $G=(\bar{x}_1 \wedge x_3) \vee (x_2 \wedge \bar{x}_3)$, the restriction of G by ρ is

$$G|_\rho \equiv (\bar{0} \wedge x_3) \vee (1 \wedge \bar{x}_3) \equiv 1 .$$

The notation $\rho|_X$ has the usual meaning of the function ρ restricted to domain X. A minterm is a minimal assignment of variables to 0, 1 that forces $G=1$; its size is the number of variables so set. In the above example there are three minterms: $x_1=0$, $x_3=1$; $x_2=1$, $x_3=0$; and $x_1=0$, $x_2=1$.

Definition 3.2. Given p we will consider the *random restriction* ρ defined by:

$$\Pr[\rho(i)=0]=\Pr[\rho(i)=1]=(1-p)/2 , \qquad \Pr[\rho(i)=*]=p ,$$

so that $G|_\rho$ becomes a random function. In application p will be small. ρ will serve to homogenize G and turn it into a simple function with high probability. The main result is purely probabilistic in statement.

Theorem 3.3. *Let* $G=G_1 \wedge G_2 \wedge \cdots \wedge G_w$, *where each*

$$G_i = y_{i1} \vee \cdots \vee y_{ia_i} ,$$

each y_{ij} *is an atom, and all* $a_i \leq t$. *Let* ρ *be the random restriction defined in* 3.2. *Let* E_s *be the event that* $G|_\rho$ *has a minterm of size at least* s. *Then*

$$\Pr[E_S] < (5pt)^s .$$

Remark. The number of conjuncts w does not affect the bound, reminiscent of the Lovász Local Lemma.

Proof. We actually show a much stronger result. Let F be any Boolean function. Then for the conditional probability we have

$$\Pr[E_S \mid F|_\rho \equiv 1] < (5pt)^s . \tag{3.4}$$

(We use the convention that if the condition is unsatisfiable then the conditional probability is considered zero.) With $F=1$ (3.4) reduces to Theorem 3.3. We prove (3.4) by induction on w. When $w=0$, $G=1$ and the result is trivial. Assume (3.4) to hold for all values less than w. We write $G=G_1 \wedge G^*$, where $G^*=G_2 \wedge \cdots \wedge G_w$, and let E_s^* be the event that G^* has a minterm of size at least s. Interchanging the roles of x_i and $\bar{x}_i$ where necessary we may, for convenience, write $G_1 = \bigvee_{i \in T} x_i$, where $|T| \leq t$. Let us assume $|T|=t$, the other cases being easier. Now either $G_1|_\rho \equiv 1$ or $G_1|_\rho \not\equiv 1$. Assuming the former, E_s holds only if E_s^* holds, so

$$\Pr[E_s \mid F|_\rho \equiv 1, G_1|_\rho \equiv 1] = \Pr[E_s^* \mid (F \wedge G_1)|_\rho \equiv 1] < (5pt)^s \tag{3.5}$$

by induction.

Now assume $G_1|_\rho \not\equiv 1$. For nonempty $Y \subset T$ and $\sigma : Y \rightarrow \{0,1\}$, $\sigma \not\equiv 0$, let $E_s^{Y,\sigma}$

be the event that $G|_\rho$ has a minterm of size at least s with assignments $x_i \leftarrow \sigma(i)$, $i \in Y$, and no other assignments with $i \in T$. Any minterm of $G|_\rho$ must force G_1 to 1 and so must include such an assignment. Thus

$$\begin{aligned}\Pr[E_s \mid F|_\rho \equiv 1, G|_\rho \not\equiv 1] &= \Pr\Big[\bigvee E_s^{Y,\sigma} \mid F|_\rho \equiv 1, G_1|_\rho \not\equiv 1\Big] \\ &\leq \sum_{Y,\sigma} \Pr[E_s^{Y,\sigma} \mid F|_\rho \equiv 1, G_1|_\rho \not\equiv 1]\,. \end{aligned} \tag{3.6}$$

There are $\binom{t}{y}$ choices of Y with y elements and $2^y - 1$ choices of σ.

Remark. $y = 1$ supplies the main term.

The conditions $F|_\rho \equiv 1$, $G_1|_\rho \not\equiv 1$ give a Bayesian distribution on $\rho|_T$. $G_1|_\rho \not\equiv 1$ means precisely that $\rho(i) \in \{0, *\}$ for all $i \in T$, so that

$$\Pr[\rho(i) = * \mid G_1|_\rho \not\equiv 1] = \frac{p}{p + (1-p)/2} = \frac{2p}{1+p}$$

for all $i \in T$ and these events are mutually independent. In particular, with $|Y| = y$,

$$\Pr[\rho(Y) = * \mid G_1|_\rho \not\equiv 1] = \left[\frac{2p}{1+p}\right]^y .$$

The further condition $F|_\rho \equiv 1$ can only serve to decrease this probability. [One proof is via the FKG inequality. We refer to the survey of Graham (1983) for a discussion of this useful result.] Letting $\rho' : \{1, \ldots, n\} - Y \rightarrow \{0, 1, *\}$ be arbitrary, we claim

$$\Pr[\rho(Y) = * \mid F|_\rho \equiv 1, G_1|_\rho \not\equiv 1, \rho|_{\{1,\ldots,n\}-Y} = \rho'] \leq \left[\frac{2p}{1+p}\right]^y .$$

Given ρ', there is a unique extension ρ with $\rho(Y) = *$. If that ρ does not satisfy the above conditions, then the conditional probability is zero. If that ρ does satisfy the conditions, then so do all extensions ρ with $\rho(i) \in \{0, *\}$ for all $i \in Y$, and so there is equality above. As this holds for all ρ',

$$\Pr[\rho(Y) = * \mid F|_\rho \equiv 1, G_1|_\rho \not\equiv 1] \leq \left[\frac{2p}{1+p}\right]^y \leq (2p)^y .$$

Remark. Given $F|_\rho \equiv 1$, $G_1|_\rho \not\equiv 1$, the probability that there exists an $i \in T$, $\rho(i) = *$, is at most $2pt$. If, to the contrary, $\rho(i) \neq *$ for all $i \in T$, then all $\rho(i) = 0$ (as $G_1|_\rho \not\equiv 1$), so $G_1|_\rho \equiv 0$, hence $G|_\rho \equiv 0$ and E_s cannot occur. With $p \ll 1/t$ this greatly limits $\Pr[E_s \mid F|_\rho \equiv 1, G_1|_\rho \not\equiv 1]$.

Let $\rho' : T \rightarrow \{0, *\}$, with $\rho'(Y) = *$, be arbitrary and condition on $\rho|_T = \rho'$. Then ρ may be considered a random restriction on $\{1, \ldots, n\} - T$. The event $F|_\rho \equiv 1$ reduces to $\bar{F}|_\rho \equiv 1$, $\bar{F}$ a function on x_i, $i \notin T$. $E_s^{Y,\sigma}$ occurs if and only if $\bar{G}$ has a minterm of size at least $s-y$, where $\tilde{G}$ is the reduction of G^* to x_i, $i \not\in T$, by ρ' and σ. Calling this event $\tilde{E}_{s-y}$,

$$\Pr[E_s^{Y,\sigma} \mid F|_\rho \equiv 1, G_1|_\rho \not\equiv 1, \rho|_T = \rho'] = \Pr[\tilde{E}_{s-y} \mid \tilde{F}|_\rho \equiv 1] \leqslant (5pt)^{s-y}$$

by induction. Any ρ with $F|_\rho \equiv 1$, $G_1|_\rho \not\equiv 1$, $\rho(Y) = *$, must have $\rho|_T = \rho'$ for some ρ' of this form. Thus

$$\Pr[E_s^{Y,\sigma} \mid F|_\rho \equiv 1, G_1|_\rho \not\equiv 1, \rho(Y) = *] \leqslant (5pt)^{s-y}$$

and so

$$\begin{aligned}
&\Pr[E_s^{Y,\sigma} \mid F|_\rho \equiv 1, G_1|_\rho \not\equiv 1] \\
&\quad = \Pr[\rho(Y) = * \mid F|_\rho \equiv 1, G_1|_\rho \not\equiv 1] \Pr[E_s^{Y,\sigma} \mid F|_\rho \equiv 1, G_1|_\rho \not\equiv 1, \rho(Y) = *] \\
&\quad \leqslant (2\rho)^y (5pt)^{s-y} .
\end{aligned}$$

Plugging in to (3.6),

$$\begin{aligned}
\Pr[E_s \mid F|_\rho \equiv 1, G_1|_\rho \not\equiv 1] &\leqslant \sum_{y=1}^{t} \binom{t}{y} (2^y - 1)(2\rho)^y (5pt)^{s-y} \\
&< (5pt)^s \sum_{y=1}^{t} \left\{ \begin{matrix} t \\ y \end{matrix} \right\} \left\{ \frac{4}{5t} \right\}^y = (5pt)^s \left\langle \left\{ 1 + \frac{4}{5t} \right\}^t - 1 \right\rangle \\
&< (5pt)^s .
\end{aligned}$$

Combining this with (3.5),

$$\Pr[E_s \mid F|_\rho \equiv 1] \leqslant (5pt)^s ,$$

completing the inductive proof of (3.4) and thus Theorem 3.3. □

We give only a very rough outline of the application of Theorem 3.3 to the complexity of the *parity function*, using the language of circuit complexity. A sequence of Boolean functions P^n on n variables is called *Borel* if for some constant k and polynomial $p(n)$ P^n is realizable by a circuit of depth k, with gates alternating between $\wedge$ and $\vee$, and maximum fan-in (i.e., number of inputs) $p(n)$. Let P_n denote the parity function, $P_n(x_1, \ldots, x_n) = [x_1 + \cdots + x_n] \bmod 2$.

Theorem 3.7 (Yao 1985). *P_n is not Borel.*

Proof. From our prejudiced vantagepoint Yao proves a Ramsey theorem. Identify Boolean functions F with two-colorings of the Hamming cube $C^n = \{0, 1\}^n$. A t-face of C^n is a set of 2^t points, all of whose coordinates but t are fixed. For all $k, t, p(n)$ it is shown that if n is sufficiently large and C^n is two-colored by a $(k, p(n))$ circuit, then there is a monochromatic t-face. As the parity function has no monochromatic 1-face it cannot be Borel.

Fix a $(k, p(n))$ circuit P. For inductive purposes hold the bottom-level fan-in to size $n^{o(1)}$ and allow the second-level (counting from the bottom) fan-in to be arbitrary. Assume the gates at the bottom level are $\vee$, otherwise use the dual form. Let G be the Boolean function achieved at the second level so it has the form of Theorem 3.3. Take a random restriction ρ with $p = n^{-\epsilon}$. The probability that G has a minterm of size $n^{o(1)}$ is exponentially small. As there are only polynomially many G, there is a restriction ρ' so that all G have all minterms small and can be rewritten as an $\vee$ of $\wedge$. Now the second and third gates are both $\vee$ and so may be combined, reducing the depth by one. Continuing, we reduce P to a level-two circuit, say $\wedge$ of $\vee$. Fixing all of the variables on one bottom fan-in we can force it, and hence P, to be false – and this gives a monochromatic face in P. □

Remark. The Borel sequences G_n are a natural analog to classical Borel sets, "countable" having been replaced by "polynomial". If we allow an arbitrary polynomial-time Turing machine for "computing" G_n, then might we not call such sequences "measurable"? And if probabilistic methods have been successful in proving "non-Borelness", then might we not hope that they will provide the key to showing certain sequences "nonmeasurable"? A fantasy to be sure, and as "P = NP?" has moved to center-stage of our mathematical consciousness a fantasy that is particularly compelling to practitioners of the probabilistic method.

4. Gems

Does the heart of mathematics lie in the building of structure or in the solving of individual problems? Not an either–or question, to be sure, but one that is particularly effective in splitting the ranks of combinatorialists. Use of an algebraic structure to explain discrete phenomena will be central to some, to others grotesque. A clever argument is beautiful to the problem-solver, a curiosity to the structuralist. The very term "combinatorial methods" has, to this author, an oxymoronic character. It is the brilliant proofs, those that expand and/or transcend the known methodologies, which express the soul of the subject. We consider three examples. Our first deals with the independence number $\alpha(G)$ of a graph.

Theorem 4.1. *Let G have vertex set $\{1, \ldots, n\}$ and let d_i denote the degree of vertex i. Then*

$$\alpha(G) \geq \sum_{i=1}^{n} 1/(d_i + 1) .$$

Proof. Let $<$ be a random ordering of the vertex set. For $1 \leq i \leq n$ let A_i be the event that all neighbors of i are greater than i in the ordering and let $C = C_<$ be the set of i for which A_i holds. As i and its neighbors are randomly ordered,

$\Pr[A_i] = 1/(d_i + 1)$ and so, by linearity of expectation,

$$E[|C_<|] = \sum_{i=1}^{n} 1/(d_i + 1) .$$

Thus for some specific ordering $|C_<| \geq \sum 1/(d_i + 1)$. Let $\{i, j\}$ be any edge of G. Either $i < j$ or $j < i$. In the first case $j \not\in C_<$ and in the second case $i \not\in C_<$. That is, $C_<$ is an independent set. □

Turán's Theorem can be shown directly from this result since, fixing the total number of edges, $\sum 1/(d_i + 1)$ is minimized when the d_i's are as nearly equal as possible.

The second gem involves the *tournament ranking function* $h(n)$ given at the end of section 1. Recall the Erdős–Moon bound

$$h(n) \leq \frac{1}{2}\binom{n}{2} + cn^{3/2}(\ln n)^{1/2} ,$$

given by an elementary use of probabilistic methods. In Spencer (1972), as part of the present author's dissertation, the lower bound

$$h(n) > \frac{1}{2}\binom{n}{2} + cn^{3/2}$$

was shown, with a probabilistic proof that we shall not discuss here. The gap of $(\ln n)^{1/2}$ was resolved by this author in 1977. Using a very complicated proof,

$$h(n) < \frac{1}{2}\binom{n}{2} + cn^{3/2}$$

(with a different constant) was shown. We present this result but with a far more ingenious argument due to de la Vega (1983). We actually show that if T is a random tournament, then almost always

$$f(T, P) \leq \frac{1}{2}\binom{n}{2} + cn^{3/2}$$

for all permutations P.

Recall that because there were $n!$ permutations in section 1 it was required that the tail distribution be less than $1/n! \sim e^{-n \ln n}$ which required being $(n \ln n)^{1/2}$ standard deviations off the mean. We are guided by the notion that had there been only K^n permutations the $(\ln n)^{1/2}$ term would not have appeared.

It will be convenient to translate to a zero mean. Define the *plusfit* of a permutation P over a set of games to be the number of games in which the lower-ranked player wins minus the number of games in which the higher-ranked player wins. For any P let $X_1(P)$ be the plusfit over the set of games between $P(i)$ and $P(j)$ where $1 \leq P(i) \leq n/2$, $n/2 < P(j) \leq n$. Then $X_1(P)$ had distribution S_m, where $m = n^2/4$ is the number of games played. Recall our large-deviation bound

$$\Pr[X_1(P) > \lambda m^{1/2}] < e^{-\lambda^2/2}\,.$$

Here, however, we do not really have $n!$ permutations P to content with. The value $X_1(P)$ is determined by the partition of the players $\{1, \ldots, n\} = U_1 \cup U_2$ into top and bottom halves, $U_1 = \{i\colon P(i) \leqslant n/2\}$, $U_2 = \{i\colon P(i) > n/2\}$. There are $\binom{n}{n/2} < 2^n$ such partitions. Set $\lambda = (2 \ln 2)^{1/2} \mathrm{n}^{1/2}$ so that $e^{-\lambda^2/2} = 2^{-n}$. Then almost always

$$X_1(P) < \lambda m^{1/2} = n^{3/2}[(2 \ln 2)^{1/2}/4^{1/2}]$$

for *all* permutations P.

Let $X_2(P)$ be the plusfit over the set of games between $P(i)$ and $P(j)$, where either $1 \leqslant P(i) \leqslant n/4$ and $n/4 < P(j) \leqslant n/2$ or $n/2 < P(i) \leqslant 3n/4$ and $3n/4 < P(j) \leqslant n$. Then $X_2(P)$ has distribution S_m, where $m = n^2/8$ is the number of games played. The value $X_2(P)$ is determined by the partition $\{1, \ldots, n\} = U_1 \cup U_2 \cup U_3 \cup U_4$ of the players into four quarters and there are less than 4^n such partitions. Set $\lambda = (2 \ln 4)^{1/2} n^{1/2}$ so that $e^{-\lambda^2/2} = 4^{-n}$. Then almost always $X_2(P) < \lambda m^{1/2} = n^{3/2}[(2 \ln 4)^{1/2}/8^{1/2}]$ for all permutations P.

Let $1 \leqslant k \leqslant \log_2 n$. A permutation P gives a partition of the players into 2^{k-1} pairs of groups of size $n2^{-k}$. Let $X_k(P)$ be the plusfit over all games "across the pairs", i.e., all games between $P(i)$ and $P(j)$ where

$$2s(n2^{-k}) < P(i) \leqslant (2s + 1)(n2^{-k}) < P(j) \leqslant (2s + 2)(n2^{-k})$$

for some integer s, $0 \leqslant s < 2^{k-1}$. The number of games $m = n^2/2^{k+1}$ goes down sharply with k. The value $X_k(P)$ is determined by the partition of the players into 2^k equal groups. There are far fewer than $(2^k)^n$ such partitions, a bound increasing sharply with k. The number of partitions winds up in a square-root log term (corresponding to the rapid decay of the tail of the normal distribution) and is dominated by the number of games, represented by a square-root term. More precisely, set $\lambda = (2 \ln 2^k)^{1/2} n^{1/2}$ so that $e^{-\lambda^2/2} = 2^{-kn}$. Then almost always

$$X_k(P) < \lambda m^{1/2} = n^{3/2}[(2 \ln 2^k)^{1/2}/(2k^{k+1})^{1/2}]$$

for all permutations P. Moreover, with a little care we can show that these events almost always occur simultaneously for all k.

Let $X(P)$ be the plusfit over all games. Almost always

$$X(P) = \sum_k X_k(P) \leqslant n^{3/2} \sum_{k=1}^{\infty} (2 \ln 2^k)^{1/2}/(2^{k+1})^{1/2} < 3.5 n^{1/2}$$

by some calculation, the convergence reflecting the dominance discussed above. Finally, the fit $f(T, P) = [X(P) + \binom{n}{2}]/2$, so that almost all T have

$$f(T, P) \leqslant \frac{1}{2}\binom{n}{2} + 1.75 n^{3/2}$$

for all $n!$ permutations P.

The final gem is due to Paul Erdős (1959). Over the past six decades Erdős has played a leading role in the rapid development of combinatorial analysis. He originated the probabilistic method with his 1947 paper on the Ramsey function $R(k, k)$ and has steadily overseen its growth. Combining a peripatetic life-style with a constant admonition to "prove and conjecture", his ideas have spread throughout the mathematical community. We have discussed the work of many other mathematicians but all of us build on the theorems and the spirit of "Uncle Paul".

Let $g(G)$, the girth of a graph, denote the length of the minimal cycle of G and let $\alpha(G)$ and $\chi(G)$ denote, as usual, the independence number and chromatic number of G, respectively.

Theorem 4.2. *For all k, t there exists a finite graph G with*

$$\chi(G) \geq k, \qquad g(G) \geq g.$$

This is a highly unintuitive result. If the girth is large one can construe no reason why the graph could not be colored with a few colors. Locally it is easy to three-color such a graph. We force the chromatic number up by global considerations. The most aesthetically pleasing aspect of this theorem is that probabilistic concepts do not enter at all into the statement, only the proof.

Proof. Let n be very large and let G be a random graph on $\{1, \ldots, n\}$ where each $\{i, j\}$ is placed in G with independent probability p. A t-set S is independent with probability $(1-p)^{\binom{t}{2}} \sim \mathrm{e}^{-pt^2/2}$. There are $\binom{n}{t} < n^t$ such sets. The probability that $\alpha(G) \geq t$ is at most $n^t \mathrm{e}^{-pt^2/2} = (n\,\mathrm{e}^{-pt/2})^t$. When $pt/2 > (1 + \mathrm{o}(1))(\ln n)$ this quantity is $\mathrm{o}(1)$ so that $\alpha(G) < t$ almost always. Critically, $\chi(G) \geq n/\alpha(G)$, since in any coloring each color class must be an independent set. Thus $\chi(G) > n/t$ almost always.

An elementary approach would be to select $t = n/k$ and $p = (2 + \mathrm{o}(1))k(\ln n)/n$ so that $\chi(G) > k$ almost always. This does not work by itself since G would then have $\binom{n}{3}p^3 > c(\ln n)^3$ triangles. An alteration is needed, two inches off the waist. Set $p = (4 + \mathrm{o}(1))k(\ln n)/n$ so that $\alpha(G) < n/2k$ almost always. Let X be the number of cycles of length at most g. For $3 \leq i \leq g$ there are less than n^i potential i-cycles, each of which is in G with probability p^i. By linearity of expectation

$$\mathrm{E}(X) = \sum_{i=3}^{g} (np)^i < c(\ln n)^g,$$

where c depends on g. With g fixed, $\mathrm{E}(X) = \mathrm{o}(n)$ so that $X < n/2$ almost always. Almost always $\alpha(G) < n/2k$ and G has fewer than $n/2$ cycles of size at most g. Fix a G with these properties. Delete from G one vertex from each such cycle. The remaining G' has no small cycles, at least $n/2$ vertices, and $\alpha(G') < n/2k$. Then $\chi(G') \geq |G'|/\alpha(G') > k$. □

Appendix: Bounding of large deviations

We give here some basic bounds on large deviations that are useful when employing the probabilistic method. Our treatment is self-contained. Most of the results may be found in, or immediately derived from, the seminal paper of Chernoff (1952). While we are guided by asymptotic considerations the inequalities are proven for all values of the parameters in the specified region. The first result, while specialized, contains basic ideas found throughout this appendix.

Theorem A.1. *Let X_i, $1 \leq i \leq n$, be mutually independent random variables with*

$$\Pr[X_i = +1] = \Pr[X_i = -1] = \tfrac{1}{2}$$

and set, following the usual convention,

$$S_n = X_1 + \cdots + X_n .$$

Let $a > 0$. Then

$$\Pr[S_n > a] < \mathrm{e}^{-a^2/2n} .$$

Remark. For large n the Central Limit Theorem implies that S_n is approximately normal with zero mean and standard deviation $\sqrt{n}$. In particular, for any fixed u

$$\lim_{n \to \infty} \Pr[S_n > u\sqrt{n}] = \int_{t=u}^{\infty} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-t^2/2} \, \mathrm{d}t ,$$

which one can show directly is less than $\mathrm{e}^{-u^2/2}$. Our proof, we emphasize once again, is valid for all n and all $a > 0$.

We require *Markov's inequality* which states: if Y is an arbitrary nonnegative random variable and $\alpha > 0$, then

$$\Pr[Y > \alpha \mathrm{E}[Y]] < 1/\alpha .$$

Proof of Theorem A.1. Fix n, a and let, for the moment, $\lambda > 0$ be arbitrary. For $1 \leq i \leq n$

$$\mathrm{E}[\mathrm{e}^{\lambda X_i}] = (\mathrm{e}^{\lambda} + \mathrm{e}^{-\lambda})/2 = \cosh(\lambda) .$$

We require the inequality

$$\cosh(\lambda) < \mathrm{e}^{\lambda^2/2} ,$$

valid for all $\lambda > 0$, the special case $\alpha = 0$ of Lemma A.5 below. (The inequality may be more easily shown by comparing the Taylor series of the two functions termwise.)

From the definition of S_n we obtain

$$e^{\lambda S_n} = \prod_{i=1}^{n} e^{\lambda X_i} .$$

Since the X_i are mutually independent so are the $e^{\lambda X_i}$. Therefore, expectations multiply and

$$E[e^{\lambda S_n}] = \prod_{i=1}^{n} E[e^{\lambda X_i}] = [\cosh(\lambda)]^n < e^{\lambda^2 n/2} .$$

We note that $S_n > a$ if and only if $e^{\lambda S_n} > e^{\lambda a}$ and apply Markov's inequality so that

$$\Pr[S_n > a] = \Pr[e^{\lambda S_n} > e^{\lambda a}] \leqslant E[e^{\lambda S_n}]/e^{\lambda a} < e^{\lambda^2 n/2 - \lambda a} .$$

We set $\lambda = a/n$ to optimize the inequality: $\Pr[S_n > a] < e^{-a^2/2n}$ as claimed. □

By symmetry we immediately have:

Corollary A.2. *Under the assumptions of Theorem A.*1,

$$\Pr[|S_n| > a] < 2\,e^{-a^2/2n} .$$

Our remaining results will deal with distributions X of the following prescribed type.

Assumptions A.3.

$$p_1, \ldots, p_n \in [0, 1] , \qquad p = (p_1 + \cdots + p_n)/n ,$$
$$X_1, \ldots, X_n \text{ mutually independent with}$$
$$\Pr[X_i = 1 - p_i] = p_i , \qquad \Pr[X_i = -p_i] = 1 - p_i ,$$
$$X = X_1 + \cdots + X_n .$$

Remark. Clearly $E[X] = E[X_i] = 0$. When all $p_i = \frac{1}{2}$, X has distribution $S_n/2$. When all $p_i = p$, X has distribution $B(n, p) - np$ where $B(n, p)$ is the usual binomial distribution.

Theorem A.4. *Under Assumptions A.*3 *and with* $a > 0$

$$\Pr[X > a] \leqslant e^{-2a^2/n} .$$

Lemma A.5. *For all reals* α, β, *with* $|\alpha| \leqslant 1$,

$$\cosh(\beta) + \alpha \sinh(\beta) \leqslant e^{\beta^2/2 + \alpha\beta} .$$

Proof. This is immediate if $\alpha = +1$, $\alpha = -1$, or $|\beta| \geqslant 100$. If the lemma were false the function

$$f(\alpha,\beta) = \cosh(\beta) + \alpha \sinh(\beta) - e^{\beta^2/2 + \alpha\beta}$$

would assume a negative global minimum in the interior of the rectangle

$$R = \{(\alpha, \beta) \colon |\alpha| \leq 1, |\beta| \leq 100\} .$$

Setting partial derivatives equal to zero we find

$$\sinh(\beta) + \alpha \cosh(\beta) = (\alpha + \beta)\, \mathrm{e}^{\beta^2/2 + \alpha\beta} ,$$
$$\sinh(\beta) = \beta\, \mathrm{e}^{\beta^2/2 + \alpha\beta}$$

and thus $\tanh(\beta) = \beta$ which implies $\beta = 0$. But $f(\alpha, 0) = 0$ for all α, giving a contradiction. □

Lemma A.6. *For all $\theta \in [0, 1]$ and all λ*

$$\theta\, \mathrm{e}^{\lambda(1-\theta)} + (1 - \theta)\, \mathrm{e}^{-\lambda\theta} \leq \mathrm{e}^{\lambda^2/8} .$$

Proof. Setting $\theta = (1 + \alpha)/2$ and $\lambda = 2\beta$, Lemma A.6 reduces to Lemma A.5. □

Proof of Theorem A.4. Let, for the moment, $\lambda > 0$ be arbitrary. Then

$$\mathrm{E}[\mathrm{e}^{\lambda X_i}] = p_i\, \mathrm{e}^{\lambda(1-p_i)} + (1 - p_i)\, \mathrm{e}^{-\lambda p_i} \leq \mathrm{e}^{\lambda^2/8} ,$$

by Lemma A.6 and

$$\mathrm{E}[\mathrm{e}^{\lambda X}] = \prod_{i=1}^{n} \mathrm{E}[\mathrm{e}^{\lambda X_i}] \leq \mathrm{e}^{\lambda^2 n/8} .$$

Applying Markov's inequality,

$$\Pr[X > a] = \Pr[\mathrm{e}^{\lambda X} > \mathrm{e}^{\lambda a}] \leq \mathrm{E}[\mathrm{e}^{\lambda X}]/\mathrm{e}^{\lambda a} \leq \mathrm{e}^{\lambda^2 n/8 - \lambda a} .$$

We set $\lambda = 4a/n$ to optimize the inequality: $\Pr[X > a] \leq \mathrm{e}^{-2a^2/n}$ as claimed. □

Again by symmetry we immediately have:

Corollary A.7. *Under Assumptions A.3 and with $a > 0$*

$$\Pr[|X| > a] \leq 2\, \mathrm{e}^{-2a^2/n} .$$

Under Assumptions A.3, with $\lambda > 0$ arbitrary,

$$\mathrm{E}[\mathrm{e}^{\lambda X}] = \prod_{i=1}^{n} \mathrm{E}[\mathrm{e}^{\lambda X_i}] = \prod_{i=1}^{n} [p_i\, \mathrm{e}^{\lambda(1-p_i)} + (1 - p_i)\, \mathrm{e}^{-\lambda p_i}]$$
$$= \mathrm{e}^{-\lambda p n} \prod_{i=1}^{n} [p_i\, \mathrm{e}^{\lambda} + (1 - p_i)] .$$

With λ fixed the function

$$f(x) = \ln[x\, \mathrm{e}^{\lambda} + 1 - x] = \ln[Bx + 1] , \quad \text{with } B = \mathrm{e}^{\lambda} - 1 ,$$

is concave and hence (Jensen's inequality)

$$\sum_{i=1}^{n} f(p_i) \leqslant nf(p) .$$

Exponentiating both sides gives

$$\prod_{i=1}^{n} [p_i \, \mathrm{e}^{\lambda} + (1 - p_i)] \leqslant [p \, \mathrm{e}^{\lambda} + (1 - p)]^n ,$$

so that

$$\mathrm{E}[\mathrm{e}^{\lambda X}] \leqslant \mathrm{e}^{-\lambda pn}[p \, \mathrm{e}^{\lambda} + (1 - p)]^n . \tag{A.8}$$

Theorem A.9. *Under Assumptions A.3 and with $a > 0$*

$$\Pr[X > a] \leqslant \mathrm{e}^{-\lambda pn}[p \, \mathrm{e}^{\lambda} + (1 - p)]^n \, \mathrm{e}^{-\lambda a}$$

for all $\lambda > 0$.

Proof. $\Pr[X > a] = \Pr[\mathrm{e}^{\lambda X} > \mathrm{e}^{\lambda a}] \leqslant \mathrm{E}[\mathrm{e}^{\lambda X}]/\mathrm{e}^{\lambda a}$. Now apply (A.8). □

Remark. For given p, n and a, an optimal assignment of λ in Theorem A.9 is found by elementary calculus to be

$$\lambda = \ln\left[\left(\frac{1 - p}{p}\right)\left(\frac{a + np}{n - (a + np)}\right)\right] .$$

This value is oftentimes too cumbersome to be useful. We employ suboptimal λ to achieve more convenient results.

Setting $\lambda = \ln[1 + a/pn]$, Theorem A.9 reduces to

$$\Pr[X > a] \leqslant \exp[a - pn \ln(1 + a/pn) - a \ln(1 + a/pn)] . \tag{A.10}$$

Theorem A.11.

$$\Pr[X > a] < \mathrm{e}^{-a^2/2pn + a^3/2(pn)^2} .$$

Proof. With $u = a/pn$ apply the inequality

$$\ln(1 + u) \geqslant u - u^2/2 ,$$

valid for all $u \geqslant 0$, to the right-hand side of (A.10). □

Remarks. (1) When all $p_i = p$, X has variance $np(1 - p)$. With $p = o(1)$ and $a = o(pn)$ this bound reflects the approximation of X by a normal distribution with variance $\sim np$.

(2) The bound of Theorem A.11 hits a minimum at $a = 2pn/3$. For $a > 2pn/3$ we have the simple bound

$$\Pr[X > a] \leqslant \Pr[X > 2pn/3] < \mathrm{e}^{-2pn/27} .$$

Theorem A.12. *For $\beta \geqslant 1$*

$$\Pr[X \geqslant (\beta - 1)pn] \leqslant [e^{\beta-1}\beta^{-\beta}]^{pn} .$$

Proof. Direct "plug in" to (A.10). □

Remark. $X + pn$ is the number of successes in n independent trials when the probability of success in the ith trial is p_i.

Theorem A.13. *Under Assumptions A.3 and with $a > 0$*

$$\Pr[X < -a] \leqslant e^{-a^2/2pn} .$$

Remark. One cannot simply employ "symmetry" as then the roles of p and $1 - p$ are interchanged.

Proof. Let $\lambda > 0$ be, for the moment, arbitrary. Then

$$\begin{aligned} E[e^{-\lambda X}] &= \prod_{i=1}^{n} E[e^{-\lambda X_i}] = \prod_{i=1}^{n} [p_i\, e^{-\lambda(1-p_i)} + (1-p_i)\, e^{\lambda p_i}] \\ &= e^{\lambda pn} \prod_{i=1}^{n} [p_i\, e^{-\lambda} + (1-p_i)] . \end{aligned}$$

With λ fixed, the function

$$f(x) = \ln[x\, e^{-\lambda} + (1-x)] = \ln[Bx + 1] , \quad \text{with } B = e^{-\lambda} - 1 ,$$

is concave. (That B is here negative is immaterial.) Thus

$$\sum_{i=1}^{n} f(p_i) \leqslant nf(p) .$$

Exponentiating both sides gives

$$E[e^{-\lambda X}] \leqslant e^{\lambda pn}[p\, e^{-\lambda} + (1-p)]^n ,$$

analogous to (A.8). Then

$$\Pr[X < -a] = \Pr[e^{-\lambda X} > e^{\lambda a}] \leqslant e^{\lambda pn}[p\, e^{-\lambda} + (1-p)]^n\, e^{-\lambda a} ,$$

analogous to Theorem A.9. We employ the inequality

$$1 + u \leqslant e^u ,$$

valid for all u, so that

$$p\, e^{-\lambda} + (1-p) = 1 + (e^{-\lambda} - 1)p \leqslant \exp[p(e^{-\lambda} - 1)]$$

and

$$\Pr[X < -a] \leqslant \exp[\lambda pn + np(e^{-\lambda} - 1) - \lambda a] = \exp[np(e^{-\lambda} - 1 + \lambda) - \lambda a] .$$

We employ the inequality

$$e^{-\lambda} \leqslant 1 - \lambda + \lambda^2/2\,,$$

valid for all $\lambda > 0$. (Note: the analogous inequality $e^{\lambda} \leqslant 1 + \lambda + \lambda^2/2$ is *not* valid for $\lambda > 0$ and so this method, when applied to $\Pr[X > a]$, requires an "error" term such as is found in Theorem A.11.) Now

$$\Pr[X < -a] \leqslant e^{np\lambda^2/2 - \lambda a}\,.$$

We set $\lambda = a/np$ to optimize the inequality: $\Pr[X < -a] \leqslant e^{-a^2/2pn}$ as claimed. □

References

Alon, N., and J. Spencer

[1992] *The Probabilistic Method* (Wiley, New York).

Beck, J.

[1978] On 3-chromatic hypergraphs, *Discrete Math.* **24**, 127–137.

Chernoff, H.

[1952] A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations, *Ann. Math. Stat.* **23**, 493–509.

de la Vega, W.F.

[1983] On the maximal cardinality of a consistent set of arcs in a random tournament, *J. Combin. Theory B* **35**, 328–332.

Erdős, P.

[1947] Some remarks on the theory of graphs, *Bull. Amer Math. Soc.* **53**, 292–294.

[1959] Graph theory and probability, *Canad. J. Math.* **11**, 34–38.

[1961] Graph theory and probability II, *Canad. J. Math.* **13**, 346–352.

[1963a] On a problem in graph theory, *Math. Gaz.* **47**, 220–223.

[1963b] On a combinatorial problem I, *Nordisk Mat. Tidskr.* **11**, 5–10.

[1964] On a combinatorial problem II, *Acta Math. Acad. Sci. Hungar.* **15**, 445–447.

Erdős, P., and H. Hanani

[1963] On a limit theorem in combinatorial analysis, *Publ. Math. Debrecen* **10**, 10–13.

Erdős, P., and L. Lovász

[1975] Problems and results on 3-chromatic hypergraphs and some related questions, in: *Finite and Infinite Sets*, eds. A. Hajnal, L. Lovász and V.T. Sós, *Colloq. Math. Soc. János Bolyai* **37**, 609–628.

Erdős, P., and J. Moon

[1965] On sets of consistent arcs in a tournament, *Canad. Math. Bull.* **8**, 269–271.

Erdős, P., and J. Spencer

[1974] *Probabilistic Methods in Combinatorics* (Academic Press/Akadémiai Kiadó, New York/Budapest).

Frankl, P., and V. Rödl

[1985] Near perfect coverings in graphs and hypergraphs, *European J. Combin.* **6**, 317–326.

Graham, R.

[1983] Applications of the FKG inequality and its relatives, in: *Mathematical Programming. The State of the Art*, eds. A. Bachem et al. (Springer, Berlin) pp. 115–131.

Hastad, J.

[1988] Almost optimal lower bounds for small depth circuits, in: *Randomness and Computation*, ed. S. Micali, *Advances in Computer Research*, Vol. 5 (JAI Press, Greenwich, CT) pp. 143–170.

Karp, R.M.

[1976] The probabilistic analysis of some combinatorial search algorithms, in: *Algorithms and Complexity: New Directions and Recent Results*, ed. J.F. Traub (Academic Press, New York) pp. 1–9.

Kleitman, D.J.

[1966] On a combinatorial conjecture of Erdős, *J. Combin. Theory* **1**, 209–214.

Mulmuley, K., U.V. Vazirani and V.V. Vazirani

[1987] Matching is as easy as matrix inversion, *Combinatorica* **7**, 105–113.

Rödl, V.

[1985] Proof of the Erdős Hanani conjecture, *European J. Combin.* **6**, 69–78.

Spencer, J.

[1972] Optimal ranking of tournaments, *Networks* **1**, 135–138.

[1977] Asymptotic lower bounds for Ramsey functions, *Discrete Math.* **20**, 69–76.

[1978] Nonconstructive methods in discrete math, in: *Studies in Combinatorics*, ed. G.C. Rota (Mathematical Association of America, New York) pp. 142–178.

[1985] Six standard deviations suffice, *Trans. Amer Math. Soc.* **289**, 679–706.

[1994] *Ten Lectures on the Probabilistic Method*, 2nd Ed. (SIAM, Philadelphia, PA).

Szele, T.

[1943] Kombinatorikai vizsgálatok az irányitott teljes gráffai kapcsolatban, *Mat. Fiz. Lapok* **50**, 223–256.

Yao, A.C.-C.

[1985] Separating the polynomial-time hierarchy by oracles, in: *Proc. 26th Annu. IEEE Symp. on Foundations of Computer Science* (IEEE Computer Society Press, Washington, DC) pp. 1–10.

CHAPTER 34

Topological Methods

A. BJÖRNER

Department of Mathematics, Royal Institute of Technology, S-10044 Stockholm, Sweden.

Contents

HANDBOOK OF COMBINATORICS
Edited by R. Graham, M. Grötschel and L. Lovász

1817

1. Introduction

In this chapter we discuss some of the ways in which topology has been used in combinatorics. The emphasis is on methods for solving genuine combinatorial problems that initially do not involve any topology – rather than on more theoretical aspects of the combinatorics–topology connection – and the selection of material reflects this aim.

The chapter is divided into two parts. In part I several examples are presented which illustrate different uses of topology in combinatorics. In part II we have gathered a number of tools which have proven useful for dealing with the topological structure found in combinatorial situations. Also, a brief review of relevant parts of combinatorial topology is given. Part II, which begins with section 9, is intended mainly for reference purposes.

Among the examples in part I one can discern at least four ways in which topology enters the combinatorial sphere. Of course, it is in the nature of such comments that no rigid demarcation lines could or should be drawn. Also other connections exist between topology and combinatorics that follow different paths.

(i) In the first three examples (sections 2–4) topology enters in the following way. First a relevant simplicial complex is identified in the combinatorial context. Then it is shown that this complex has sufficiently favorable properties to allow application of some theorem of algebraic topology, which implies the combinatorial conclusion.

(ii) A different approach is seen in section 5 and in Bárány's proof in section 4. There a combinatorial configuration is represented in concrete fashion in $\mathbb{R}^d$ or on the d-sphere, and a topological result (Borsuk's Theorem) has the desired effect on the configuration.

(iii) The case of oriented matroids (section 7) is unique. For these combinatorial objects there is a *topological representation theorem*, saying that oriented matroids are the same thing as arrangements of certain codimension one subspheres in a sphere. Of course, in this situation the topological perspective is always at hand as an alternative way of looking at these objects. Some nontrivial properties of oriented matroids find particularly simple proofs in this way.

(iv) The need for homotopy results in combinatorics sometimes arises as follows. Say we want to define some property $\mathcal{P}$ at all vertices of a connected graph $G = (V, E)$. We start by defining $\mathcal{P}$ at some root node r, and then give a rule for how to define $\mathcal{P}$ at v's neighbors, having already defined it at $v \in V$. The problem of consistency arises: Can different paths from r to v lead to different definitions of $\mathcal{P}$ at v? One strategy for dealing with this is to define "elementary homotopies", meaning certain pairs of paths which can be exchanged without affecting the result (usually such pairs form small circuits such as triangles and squares). Then we need a "homotopy theorem" saying that any path from r to v can be deformed into any other such path using elementary homotopies. Tutte's and Maurer's homotopy theorems (section 6) are of this kind. From a topological point of view, the "elementary homotopies" mean that certain 2-cells are attached to the graph, and the homotopy theorem then says that the resulting 2-complex is simply connected.

Being topologically k-connected has a direct combinatorial meaning for $k = 0$ (connected), and, as we have seen, also for $k = 1$ (simply connected). The way that higher connectivity influences combinatorics is more subtle; see the examples in sections 4 and 6.

In section 8 a glimpse is given of the Hard Lefschetz Theorem and its applications to combinatorics found by R. Stanley. The question here is of finding a complex projective variety whose topology (in the form of its cohomology ring) is relevant to the combinatorics at hand. This rarefied method has found a few striking applications. Since it deals more with algebraic-geometric matters (the topology is somewhat subordinate), section 8 is rather loosely connected with the rest of the chapter.

Topological reasoning plays an important role in connection with several other topics in discrete mathematics not treated here. Among these, let us mention: embeddings of graphs in surfaces (see chapter 5 by Thomassen), convex polytopes (see chapter 18 by Klee and Kleinschmidt and also Bayer and Lee 1993), arrangements of subspaces (see Orlik and Terao 1992 and Björner 1994a), group-related incidence geometries (diagram geometries, chamber systems, posets of subgroups) (see Buekenhout 1995, Ronan 1989 and Webb 1987), computational geometry and realization spaces (see Bokowski and Sturmfels 1989), lower bounds for decision and computation trees (see chapter 32 by Alon and also Björner 1994a).

Notation and terminology is explained in part II. We treat simplicial complexes and posets almost interchangeably. The order complex of a poset and the poset of faces of a complex – these two constructions take posets to complexes and vice versa, and no ambiguity can arise from the topological point of view.

This chapter was written in 1988, and was revised and updated in 1989 and 1993.

PART I. EXAMPLES

2. Evasive graph properties

By a *graph property* we shall understand a property of graphs which is isomorphism-invariant: if $G_1 \cong G_2$ then G_1 has the property if and only if G_2 does. The following discussion will concern simple graphs having some fixed vertex set V. These graphs can be identified with the various subsets of $\binom{V}{2}$. Also, it is convenient to identify a graph property with the subset of the power set $\mathbf{2}^{\binom{V}{2}}$ which consists of all graphs having the property. A graph property $\mathcal{P} \subseteq \mathbf{2}^{\binom{V}{2}}$ is called *monotone* if it is preserved under deletion of edges. It is called *trivial* if either $\mathcal{P} = \emptyset$ or $\mathcal{P} = \mathbf{2}^{\binom{V}{2}}$.

In section 4.5 of chapter 23 by Bollobás the concept of *complexity* (sometimes called "argument complexity") of graph properties is discussed. Also, *evasive* graph properties are defined as those of maximal complexity. The following result (stated as Theorem 4.5.5 in chapter 23) confirms for prime-power number of vertices n a well-known conjecture.

Theorem 2.1 (Kahn, Saks and Sturtevant 1984). *Let $n = p^k$ where p is a prime. Then every non-trivial monotone property of graphs with n vertices is evasive.*

We will sketch the proof of Kahn et al. to show the way in which topology is used.

Suppose that card $V = p^k$, p prime, and that $\mathcal{P} \neq \emptyset$ is a monotone nonevasive graph property. $\mathcal{P}$ is a family of subsets of $\binom{V}{2}$ closed under the formation of subsets – i.e., a simplicial complex. The conclusion we want to draw is that $\mathcal{P}$ is trivial, which, since $\mathcal{P} \neq \emptyset$, must mean that $\binom{V}{2} \in \mathcal{P}$ – i.e., topologically $\mathcal{P}$ is the full simplex.

These two facts are crucial:

2.2. The geometric realization $\|\mathcal{P}\|$ is contractible.

2.3. There exists a group Γ of simplicial automorphisms of $\mathcal{P}$ which acts transitively on $\binom{V}{2}$ and which has a normal p-subgroup Γ_1, such that Γ/Γ_1 is cyclic.

For (2.2) one argues that the monotone property $\mathcal{P}$ is not evasive in the algorithmic sense defined above if and only if as a simplicial complex $\mathcal{P}$ is nonevasive in the recursive sense of (11.1). By (11.1) all nonevasive complexes are contractible.

The group Γ needed in (2.3) is constructed as follows. Identify V with the finite field $\mathrm{GF}(p^k)$. Let $\Gamma = \{x \longmapsto ax + b \mid a, b \in \mathrm{GF}(p^k), a \neq 0\}$ and $\Gamma_1 = \{x \longmapsto x + b \mid b \in \mathrm{GF}(p^k)\}$. The assumption that $\mathcal{P}$ is an isomorphism-invariant property of graphs on V means that if γ is any permutation of V – in particular, if $\gamma \in \Gamma$ – then $A \in \mathcal{P}$ if and only if $\gamma(A) \in \mathcal{P}$. Hence, Γ is a group of simplicial automorphisms of $\mathcal{P}$. One checks that Γ is doubly transitive on $V = \mathrm{GF}(p^k)$ and that the subgroup Γ_1 has the required properties.

By a theorem of Oliver (1975), any action of a finite group Γ, having a subgroup Γ_1 with the stated properties, on a finite $\mathbb{Z}_p$-acyclic simplicial complex must have stationary points. Since our complex $\mathcal{P}$ is $\mathbb{Z}_p$-acyclic (being contractible), this means that there exists some point $x \in \|\mathcal{P}\|$ such that $\gamma(x) = x$ for all $\gamma \in \Gamma$. The point x is carried by the relative interior of a unique face $G \in \mathcal{P}$ (the lowest-dimensional face containing it), and the fact that x is stationary implies that $\gamma(G) = G$ for all $\gamma \in \Gamma$. But since Γ is transitive on $\binom{V}{2}$ this is impossible unless $G = \binom{V}{2}$. Hence, $\binom{V}{2} \in \mathcal{P}$, and we are done.

It has been conjectured that *all* non-trivial monotone graph properties are evasive. This conjecture remains open for all non-prime-power $n \geqslant 10$; the $n = 6$ case was verified by Kahn et al. (1984). The evasiveness conjecture has been proven also for the case of bipartite graphs by Yao (1988), using the topological method.

3. Fixed points in posets

A poset P is said to have the *fixed-point property* if every order-preserving self-map $f: P \rightarrow P$ has a fixed point $x = f(x)$. It was shown by A. C. Davis and A. Tarski that a lattice has the fixed-point property if and only if it is complete (meaning that meets and joins exist for subsets of arbitrary cardinality). It has long been an open problem to find some characterization of the finite posets which have the fixed-point property. See Rival (1985) for references to work in this area. In the absence of such a characterization efforts have been directed toward finding nontrivial classes of finite posets which have the fixed point property. For this the Lefschetz fixed-point theorem has proved to be useful.

Let L be a finite lattice and $z \in L$. Then y is said to be a *complement* of z, written $y \perp z$, if $y \wedge z = \hat{0}$ and $y \vee z = \hat{1}$. Let $\mathscr{C}\mathrm{o}(z) = \{y \in L \mid y \perp z\}$. The lattice L is called *complemented* if $\mathscr{C}\mathrm{o}(z) \neq \emptyset$ for all $z \in L$.

A finite lattice L has the fixed point property, as is easy to see. It is more interesting to look at the *proper part* $\bar{L} = L - \{\hat{0}, \hat{1}\}$ of the lattice, which may or may not have the fixed point property. This is also natural from the point of view of lattice automorphisms, for which every nontrivial fixed point must lie in $\bar{L}$.

Theorem 3.1 (Baclawski and Björner 1979, 1981). *Let L be a finite lattice and $z \in \bar{L}$. Then the poset $\bar{L} - \mathscr{C}\mathrm{o}(z)$ has the fixed point property. In particular, if L is noncomplemented then $\bar{L}$ has the fixed point property.*

By Theorem 10.15 the order complex $\Delta(\bar{L} - \mathscr{C}\mathrm{o}(z))$ is contractible, and therefore by Lefschetz's Theorem 13.4 it has the *topological* fixed point property. From this the result easily follows.

For example, let L be a finite Boolean lattice of order n. Then $\bar{L}$ has $(n-1)!$ fixed-point-free automorphisms, but the removal of any one element from $\bar{L}$ leads to a poset with the fixed point property.

The preceding argument is, of course, applicable to *any* $\mathbb{Q}$-acyclic finite poset [see (11.1) for some other combinatorially defined classes of such]. Also, with this method one can prove more about the combinatorial structure of the *fixed-point sets* $P^f = \{x \in P \mid x = f(x)\}$ than merely that they are nonempty.

Let $f: P \rightarrow P$ be an order-preserving mapping of a finite $\mathbb{Q}$-acyclic poset. Then the Möbius function μ computed over P^f augmented by new bottom and top elements must equal zero: $\mu(P^f) = 0$. This follows from the Hopf trace formula, see (13.5) and the comments following it. A consequence is that for instance two or more incomparable points cannot alone form a fixed-point set in an acyclic poset. For other finite posets with the fixed point property such fixed-point sets are, however, possible.

Similarly, let $g: P \rightarrow P$ be an order-reversing mapping of a finite $\mathbb{Q}$-acyclic poset. Then the Hopf trace formula (13.2) specializes to $\mu(P_g) = 0$, where $P_g = \{x \in P \mid x = g^2(x) \leqslant g(x)\}$. In particular, if no $x \in P$ satifies $x = g^2(x) < g(x)$ then g has a *unique* fixed point. See Baclawski and Björner (1979) for further details and examples.

4. Kneser's Conjecture

Consider the collection of all n-element subsets of a $(2n+k)$-element set, $n \geqslant 1, k \geqslant 0$. It is easy to partition this collection into $k+2$ classes so that every pair of n-sets within the same class has nonempty intersection. Can the same be done with only $k+1$ classes? M. Kneser conjectured in 1955 that the answer is negative, and this was later confirmed by L. Lovász.

Theorem 4.1 (Lovász 1978). *If the n-subsets of a $(2n+k)$-element set are partitioned into $k+1$ classes, then some class will contain a pair of disjoint n-sets.*

Lovász's proof relies on Borsuk's Theorem 13.6 and homotopical connectivity arguments. Soon after Lovász's breakthrough a simpler way of deducing Kneser's Conjecture from Borsuk's Theorem was discovered by Bárány (1978). However, Lovász's proof method is applicable also to other situations and hence perhaps of greater general interest. See also chapter 24 by Frankl for a discussion of this result.

Let us first sketch Bárány's proof. By a theorem of Gale (1956) (see also Schrijver 1978), for $n, k \geqslant 1$ there exist $2n+k$ points on the sphere $\boldsymbol{S}^k$ such that any open hemisphere contains at least n of them. Partition the n-subsets of these points into classes $\mathscr{C}_0, \mathscr{C}_1, \ldots, \mathscr{C}_k$. For $0 \leqslant i \leqslant k$, let $\mathscr{O}_i$ be the set of all points $x \in \boldsymbol{S}^k$ such that the open hemisphere around x contains an n-subset from the class $\mathscr{C}_i$. Then $(\mathscr{O}_i)_{0 \leqslant i \leqslant k}$ gives a covering of $\boldsymbol{S}^k$ by open sets. Part (i) of Borsuk's Theorem 13.6 implies that one of the sets, say $\mathscr{O}_k$, contains antipodal points. But the open hemispheres around these points are disjoint and both contain n-subsets from the class $\mathscr{C}_k$. Hence, $\mathscr{C}_k$ contains a pair of disjoint n-sets.

For Lovász's proof it is best to think of the problem in graph-theoretic terms. Define a graph $\mathrm{KG}_{n,k}$ as follows: The vertices are the n-subsets of some fixed $(2n+k)$-element set X and the edges are formed by the pairs of disjoint n-sets. Then Theorem 4.1 can be reformulated: *The Kneser graph* $\mathrm{KG}_{n,k}$ *is not* $(k+1)$-*colorable.*

For any graph $G=(V,E)$ let $\mathscr{N}(G)$ denote the simplicial complex, called the *neighborhood complex*, whose vertex set is V and whose simplices are those sets of vertices which have a common neighbor (i.e., $A \in \mathscr{N}(G)$ iff there exists $v \in V$ such that $\{v,a\} \in E$ for all $a \in A$). The topology of this complex has surprising combinatorial content.

Theorem 4.2 (Lovász 1978). *For any finite graph G, if $\mathscr{N}(G)$ is $(k-1)$-connected, then G is not $(k+1)$-colorable.*

To prove Theorem 4.1 it will then suffice to show that $\mathscr{N}(\mathrm{KG}_{n,k})$ is $(k-1)$-connected. This can be done as follows. Let $P=\{A \subseteq X \mid n \leqslant \operatorname{card} A \leqslant n+k\}$. Ordered by containment P is a subposet of the Boolean lattice $B(X)$ of all subsets of X. $B(X)$ is shellable (11.10) (iv), hence by (11.2) and Theorem 11.14 P is $(k-1)$-connected. Let C be the crosscut of n-element sets. By Theorem 10.8, P and the

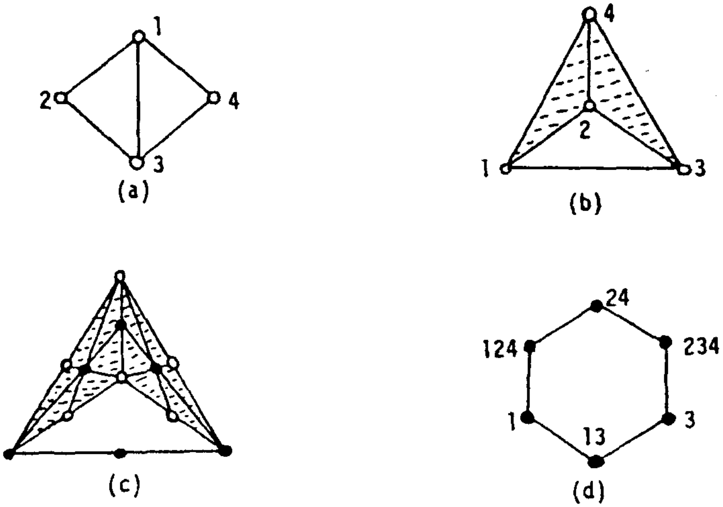


Figure 1.

crosscut complex $\Gamma(P,C)$ are homotopy equivalent. It follows that $\Gamma(P,C)$, which is the same thing as $\mathcal{N}(\mathrm{KG}_{n,k})$, is also $(k-1)$-connected.

The known proofs for Theorem 4.2 are more involved. A very elegant functorial argument was given by Walker (1983a), which we will sketch here in briefest possible fashion. The same general argument was also found by Lovász (unpublished lecture notes) as a variation of his original proof.

Let $G=(V,E)$ be a finite graph. The mapping $\nu:\mathcal{N}(G)\to\mathcal{N}(G)$ defined by $\nu(A)=\{v\in V \mid \{v,a\}\in E \text{ for all } a\in A\}$ has the properties

$$\text{(i)}\quad A\subseteq B \text{ implies } \nu(A)\supseteq\nu(B),\qquad \text{and (ii) } \nu^2(A)\supseteq A.$$

Let $\tilde{\mathcal{N}}(G)$ denote the order complex of the poset of fixed points of ν^2 ordered by containment. Thus, $\tilde{\mathcal{N}}(G)$ is a subcomplex of the barycentric subdivision of $\mathcal{N}(G)$. In fact, the subspace $\|\tilde{\mathcal{N}}(G)\|$ is by Corollary 10.12 a strong deformation retract of $\|\mathcal{N}(G)\|$, so $\tilde{\mathcal{N}}(G)$ and $\mathcal{N}(G)$ are of the same homotopy type. This construction is illustrated in fig. 1, where part (a) shows a graph G, (b) the neighborhood complex $\mathcal{N}(G)$, (c) its barycentric subdivision, and (d) the retract complex $\tilde{\mathcal{N}}(G)$.

Property (i) of the mapping $\nu:\mathcal{N}(G)\to\mathcal{N}(G)$ shows that ν restricts to a simplicial mapping $\nu:\tilde{\mathcal{N}}(G)\to\tilde{\mathcal{N}}(G)$, and from property (ii) it follows that $\nu^2=$ identity. Hence, $(\tilde{\mathcal{N}}(G),\nu)$ (or, to be precise, $(\|\tilde{\mathcal{N}}(G)\|,\|\nu\|)$) is an antipodality space. Furthermore, it can be shown that every graph map (mapping of the nodes which takes edges to edges) $g:G_1\to G_2$ induces an equivariant map $\tilde{g}:\tilde{\mathcal{N}}(G_1)\to\tilde{\mathcal{N}}(G_2)$. As these facts suggest, the construction $\tilde{\mathcal{N}}(\cdot)$ sets up a functor from the category of finite graphs and graph maps to the category of antipodality spaces and homotopy classes of equivariant maps, see Walker (1983a). For the example illustrated in fig. 1(d), the induced antipodal mapping of $\tilde{\mathcal{N}}(G)$ coincides with its antipodal map $x \mapsto -x$ as a circle.

For K_{k+1}, the complete graph on $k+1$ vertices, one sees that $\tilde{\mathcal{N}}(K_{k+1}) = \mathcal{N}(K_{k+1})$ is combinatorially the barycentric subdivision of the boundary of a k-simplex. It is also easy to verify that, as an antipodality space, $(\tilde{\mathcal{N}}(K_{k+1}), \nu)$ is isomorphic to the sphere (S^{k-1}, α) with its standard antipodality map $\alpha(x) = -x$.

We now have all the ingredients for a proof of Theorem 4.2. Suppose that a graph G is $(k+1)$-colorable. This is clearly equivalent to the existence of a graph map $G \to K_{k+1}$. Hence, we deduce the existence of an equivariant map $\tilde{\mathcal{N}}(G) \to \tilde{\mathcal{N}}(K_{k+1}) \cong S^{k-1}$. So by part (v) of Borsuk's Theorem 13.6, we conclude that $\tilde{\mathcal{N}}(G)$, and hence $\mathcal{N}(G)$, is not $(k-1)$-connected.

Schrijver (1978) has shown, using Bárány's method, that the conclusion of Theorem 4.1 remains true for the class of n-subsets that contain no consecutive elements $i, i+1$ in circular order (mod $2n+k$), and that this class is minimal with this property. A different application of Theorem 4.2 is given in Lovász (1983).

The following generalized "Kneser" conjecture was made by P. Erdős in 1973 and has recently been proved.

Theorem 4.3 (Alon, Frankl and Lovász 1986). *Let $n, t \geqslant 1$ and $k \geqslant 0$. If the n-subsets of a $(tn + (t-1)k)$-element set are partitioned into $k+1$ classes, then some class will contain t pairwise disjoint n-sets.*

The proof is analogous to Lovász's proof of Theorem 4.1. For general t-uniform hypergraphs H a suitable *neighborhood complex* $\mathscr{C}(H)$ is defined. It is shown that if t is a prime and $\mathscr{C}(H)$ is $(k(t-1)-1)$-connected then H is not $(k+1)$-colorable. To prove this for odd primes t the Bárány–Shlosman–Szűcs Theorem 13.8 is used rather than Borsuk's Theorem. It can be shown by an elementary argument that if Theorem 4.3 is valid for two values of t then it is also valid for their product. Hence one may assume that t is prime. See Alon et al. (1986) for the details.

Theorem 4.3 has been further generalized by Sarkaria (1990) to involve "j-wise disjoint" instead of "pairwise disjoint" families of n-sets. The proof uses a generalized Borsuk–Ulam theorem and the deleted join construction for simplicial complexes (defined in section 9).

5. Discrete applications of Borsuk's Theorem

One of the most famous consequences of Borsuk's Theorem 13.6 is undoubtedly the Ham Sandwich Theorem 13.7. This result, or some version of the "ham sandwich" argument which leads to it (outlined in connection with Theorem 13.7), can be used in certain combinatorial situations to prove that composite configurations can be split in a balanced way. Two examples of this, due to N. Alon and coauthors, will be given in this section. Also, we discuss how Borsuk's Theorem and its generalizations have been used in connection with results of "Tverberg" type. For other applications of Borsuk's Theorem to combinatorics, see Bárány and Lovász (1982), Yao and Yao (1985), and section 4. Surveys of this topic are given by Alon (1988), Bárány (1993) and Bogatyi (1986).

Suppose that $2n$ points are given in general position in the plane $\mathbb{R}^2$, half colored red and the other half blue. It is an elementary problem to show that the red points can be connected to the blue points by n nonintersecting straight-line segments. A quick argument goes like this. Of the $n!$ ways to match the blue and red points using straight-line segments, choose one which minimizes the sum of the lengths. If two of its lines intersect, they could be replaced by the sides of the quadrilateral that they span, and a new matching of even shorter length would result. No such elementary proof is known for the following generalization to higher dimensions.

Theorem 5.1 (Akiyama and Alon 1989). *Let A be a set of $d \cdot n$ points in general position (no more than d points on any hyperplane) in $\mathbb{R}^d$. Let $A = A_1 \cup A_2 \cup \cdots \cup A_d$ be a partition of A into d pairwise disjoint sets of size n. Then there exist n pairwise disjoint $(d-1)$-dimensional simplices, such that each simplex intersects each set A_i in one of its vertices, $1 \leqslant i \leqslant d$.*

The idea of Akiyama and Alon is to surround each point $p \in A$ by a small ball of radius ε, where ε is small enough that no hyperplane intersects more than d such balls. Give each ball a uniform mass distribution of measure $1/n$. Then each color class A_i, $1 \leqslant i \leqslant d$, is naturally associated with its n balls, forming a measurable set of measure 1. By the Ham Sandwich Theorem 13.7 there exists a hyperplane H which simultaneously bisects each color class. If n is odd, then H must intersect at least one ball from each A_i. General position immediately implies that H must intersect precisely one ball from each A_i, and in fact bisect this ball. By induction on n, the points on each side of H can now be assembled into disjoint simplices, and finally the points in H form one more such simplex. The argument if n is even is similar, but in that case H might have to be slightly moved to divide the points correctly for the induction step.

The next example has a more "applied" flavor. Suppose that k thieves steal a necklace with $k \cdot n$ jewels. There are t kinds of jewels on it, with $k \cdot a_i$ jewels of type $i, 1 \leqslant i \leqslant t$. The thieves want to divide the necklace fairly between them, wasting as little as possible of the precious metal in the links between jewels. They need to know in how many places they must cut the necklace? If the jewels of each kind appear contiguously on the opened necklace, then at least $t(k-1)$ cuts must be made. This number of cuts in fact always suffices. (Of course, what the thieves *really* need is a fast algorithm for where to place these cuts.)

Theorem 5.2 (Alon and West 1986, Alon 1987). *Every open necklace with $k \cdot a_i$ beads of color $i, 1 \leqslant i \leqslant t$, can be cut in at most $t(k-1)$ places so that the resulting segments can be arranged into k piles with exactly a_i beads of color i in each pile, $1 \leqslant i \leqslant t$.*

The idea for the proof is to turn the situation into a continuous problem by placing the open necklace (scaled to length 1) on the unit interval, and then to use a "ham-sandwich"-type argument there. For $k = 2$ this was done in Alon and West (1986) using Borsuk's Theorem. The extension to general k was achieved in Alon (1987) using the Bárány–Shlosman–Szűcs Theorem 13.8.

Radon's Theorem, a well-known result in convexity theory, says that any collection of $d+2$ points in $\mathbb{R}^d$ can be split into two nonempty blocks whose convex hulls have nonempty intersection. This was generalized by Tverberg (1966) as follows: *For all $p \geqslant 2$ and $d \geqslant 1$, any set of $(p-1)(d+1)+1$ points in $\mathbb{R}^d$ can be partitioned into p blocks $B_1, \ldots, B_p$ so that* $\operatorname{conv}(B_1) \cap \cdots \cap \operatorname{conv}(B_p) \neq \emptyset$. For a quite short proof of Tverberg's Theorem, see Sarkaria (1992). Results of the Radon–Tverberg type have generated a lot of interest, and recent work shows that in many cases such results rely on topological foundations that lead to formulations more general than the original ones in terms of convexity. See Eckhoff (1979) and Bárány (1993) for surveys of results of this kind.

Radon's theorem can be obtained as a consequence of Borsuk's Theorem, as was shown by Bajmóczy and Bárány (1979). Here is the connection. Let Δ^d denote the d-dimensional simplex. Bajmóczy and Bárány prove that there exists a continuous map $g : S^d \to \Delta^{d+1}$ such that the supports of $g(x)$ and $g(-x)$ are disjoint for every $x \in S^d$. Suppose now that Radon's Theorem is false; say it fails for the points $y_1, \ldots, y_{d+2}$ in $\mathbb{R}^d$. Define $f : \Delta^{d+1} \to \mathbb{R}^d$ by sending the ith vertex of Δ^{d+1} to y_i and extending linearly. Then the map $f \circ g : S^d \to \mathbb{R}^d$ would violate the Borsuk–Ulam Theorem 13.6 (ii).

In the preceding argument the map f could as well be an arbitrary continuous map (i.e., not necessarily linear). In a similar way, using Theorem 13.8 instead of Borsuk's Theorem, Bárány, Shlosman and Szűcs (1981) proved the following "topological Tverberg theorem": *Suppose that $f : \Delta^N \to \mathbb{R}^d$ is a continuous mapping, where $N = (p-1)(d+1)$ and p is prime. Then there exist p pairwise disjoint faces $\sigma_1, \ldots, \sigma_p$ of Δ^N such that $f(\sigma_1) \cap \cdots \cap f(\sigma_p) \neq \emptyset$.* It is still unknown whether the restriction to prime p is needed here in the non-linear case. See Sarkaria (1991b) for even more general results of this kind.

The following result has the general flavor of Tverberg's Theorem, and goes in an opposite direction from Theorem 5.1.

Theorem 5.3 (Živaljević and Vrećica 1992). *Let $A = A_1 \cup A_2 \cup \cdots \cup A_{d+1}$ be a set of points in $\mathbb{R}^d$ partitioned into $d+1$ pairwise disjoint sets (color classes) of size $|A_i| \geqslant 4n-1$. Then there exist n pairwise disjoint $(d+1)$-subsets $B_1, \ldots, B_n$ of A such that $|A_i \cap B_j| = 1$ for all i, j and* $\operatorname{conv}(B_1) \cap \cdots \cap \operatorname{conv}(B_n) \neq \emptyset$.

The proof for this "colored Tverberg theorem" uses a Borsuk–Ulam-type result for free $\mathbb{Z}_p$-actions, p prime, which establishes the non-existence of an equivariant map from a certain "configuration space" of sufficiently high connectivity to a sphere of appropriate dimension.

It has been conjectured by Bárány and Larman that $|A_i| \geqslant n$ suffices in Theorem 5.3. This has been proven for $d = 2$ by Bárány and Larman and for $n = 2$ by Lovász, whose proof uses Borsuk's theorem. See Živaljević and Vrećica (1992) for these references and for a fuller discussion of the status of this "colored Tverberg problem".

6. Matroids and greedoids

This section and the next are devoted to certain topological aspects of matroids and of two related structures – oriented matroids and greedoids. For the basic definitions see chapter 9 by Welsh. Additional topological facts about matroid complexes and geometric lattices are mentioned in (11.10); see also Björner (1992).

Basis complexes and partitions of graphs

The following result was proven by E. Győry and L. Lovász in response to a conjecture by A. Frank and S. Maurer.

Theorem 6.1 (Lovász 1977, Győry 1978). *Let $G = (V, E)$ be a k-connected graph, $\{v_1, v_2, \ldots, v_k\}$ a set of k vertices, and $n_1, n_2, \ldots, n_k$ positive integers with $n_1 + n_2 + \cdots + n_k = |V|$. Then there exists a partition $\{V_1, V_2, \ldots, V_k\}$ of V such that $v_i \in V_i, |V_i| = n_i$ and V_i spans a connected subgraph of $G, i = 1, 2, \ldots, k$.*

The proof of Lovász uses topological methods, that of Győry does not. At the end of this section Lovász's proof will be outlined for the case $k = 3$ in order to illustrate its use of topological reasoning. It relies on the connectivity of a certain polyhedral complex associated with certain forests in G. Similar complexes can be defined over the bases of a matroid, and more generally over the bases of a greedoid. The greedoid formulation contains the others as special cases, and we shall use it to develop the general result.We begin by recalling the definition.

A set system $(E, \mathcal{F}), \mathcal{F} \subseteq 2^E$, is called a *greedoid* if the following axioms are satisfied:

(G1) $\emptyset \in \mathcal{F}$,

(G2) for all nonempty $A \in \mathcal{F}$ there exists an $x \in A$ such that $A - x \in \mathcal{F}$,

(G3) if $A, B \in \mathcal{F}$ and $|A| > |B|$, then there exists an $x \in A - B$ such that $B \cup x \in \mathcal{F}$.

If also the extra condition (G4) is satisfied, then $(E, \mathcal{F})$ is called an *interval greedoid*:

(G4) if $A \subset B \subset C$ where $A, B, C \in \mathcal{F}$ and $A \cup x, C \cup x \in \mathcal{F}$ for some $x \in E - C$, then also $B \cup x \in \mathcal{F}$.

The sets in $\mathcal{F}$ are called *feasible* and the maximal feasible sets *bases*. All bases have the same cardinality r, which is the *rank* of the greedoid.

The only examples which will be of concern here are *matroids* (feasible sets = independent sets) and *branching greedoids* of rooted graphs (feasible sets = edge sets which form a tree containing the root node). Both are interval greedoids. For other examples and further information about greedoids, see chapter 9 by Welsh and the expository accounts Korte, Lovász and Schrader (1991) and Björner and Ziegler (1992).

The feasible sets of a greedoid do not form a simplicial complex other than in the matroid case. However, a useful topology is given by (the order complex of) the poset $\bar{\mathcal{F}} = \mathcal{F} - \{\emptyset\}$, ordered by inclusion. A greedoid $(E, \mathcal{F})$ is called *k-connected* if for each $A \in \mathcal{F}$ there exists $B \in \mathcal{F}$ with $A \subseteq B, |B - A| = \min(k, r - |A|)$ and such that $C \in \mathcal{F}$ for every $A \subseteq C \subseteq B$. Matroids are r-connected, and the branching greedoid of a k-connected rooted graph is k-connected.

Proposition 6.2 (Björner, Korte and Lovász 1985). *Let $(E, \mathcal{F})$ be a k-connected interval greedoid $(k \geqslant 2)$. Then the poset of feasible sets $(\bar{\mathcal{F}}, \subseteq)$ is (topologically) $(k-2)$-connected.*

This result follows from (11.10) (iii) via Theorem 10.8, since for the crosscut C of minimal elements in $\bar{\mathcal{F}}$ the crosscut complex $\Gamma(\bar{\mathcal{F}}, C)$ is a matroid complex of rank $\geqslant k$.

Let $\mathcal{B}$ be the collection of all bases in a greedoid $(E, \mathcal{F})$ of rank r. Two bases B_1 and B_2 are *adjacent* if $B_1 \cap B_2 \in \mathcal{F}$ and $|B_1 \cap B_2| = r - 1$. Attaching edges between all adjacent pairs we get a graph with vertex set $\mathcal{B}$, the *basis graph*.

The shortest circuits in the basis graph can be explicitly described. There are two kinds of triangles and one kind of square (quadrilateral):

6.3. Three bases $A \cup x, A \cup y, A \cup z$, where $A \in \mathcal{F}, |A| = r - 1$, span a triangle of the first kind.

6.4. Three bases $A \cup x \cup y, A \cup x \cup z, A \cup y \cup z$, where $A \in \mathcal{F}, |A| = r - 2$, span a triangle of the second kind.

6.5. Four bases $A \cup x \cup u, A \cup x \cup v, A \cup y \cup u, A \cup y \cup v$, where $A \in \mathcal{F}, |A| = r - 2$, span a square.

For branching greedoids triangles of the second kind cannot occur.

Now, attach a 2-cell (a "membrane") into each triangle and square. This gives a 2-dimensional regular cell complex $\mathcal{K}$, which we call the *basis complex*.

It is a straightforward combinatorial exercise to check that the basis complex of any 2-connected greedoid of rank $\leqslant 2$ is 1-connected (i.e., connected and simply connected). For rank 2 (the only non-trivial case) this follows directly from the exchange axiom (G3). In higher ranks the following is true.

Theorem 6.6 (Björner, Korte and Lovász 1985). *The basis complex $\mathcal{K}$ of any 3-connected interval greedoid is 1-connected.*

In order to illustrate some of the tools given in part II, we give a short proof of this. Let P be the poset of closed cells of $\mathcal{K}$ ordered by inclusion, and let Q be the top three levels of $(\mathcal{F}, \subseteq)$, i.e., the feasible sets of ranks $r-2$, $r-1$ and r. Let $f: P \to Q$ be the order-reversing map which sends each cell τ to the intersection of the bases which span τ. By Proposition 6.2 and Lemma 11.12 the poset Q is 1-connected, so by Theorem 10.5 we only have to check that the fibers $f^{-1}(Q_{\geqslant A})$

are 1-connected for all $A \in Q$. But if $r(A) = r - i$, $i = 0, 1, 2$, then $f^{-1}(Q_{\geq A})$ is the basis complex of the rank i greedoid obtained by contracting A, and we have already checked that basis complexes of rank $\leqslant 2$ greedoids are 1-connected.

Let $P = B_1 B_2 \cdots B_d$ and $Q = B_d B_{d+1} \cdots B_g$ be paths in the basis graph of a matroid, and let $PQ = B_1 B_2 \cdots B_d B_{d+1} \cdots B_g$ be their concatenation. Say that paths PQ and PRQ differ by an *elementary homotopy* if R is of the form $BCB, BCDB$ or $BCDEB$ with $B = B_d$.

Theorem 6.7 (Maurer 1973). *Let P and P' be any two paths with the same endpoints in the basis graph of a matroid. Then P can be transformed into P' via a sequence of elementary homotopies.*

Maurer's "Homotopy Theorem" 6.7 is clearly a combinatorial reformulation of Theorem 6.6 in the matroid case. An application to oriented matroids will be given in the next section.

The time has come to return to Theorem 6.1. The following outline of the proof for the $k = 3$ case is quoted from Lovász (1979) (with some adjustments in square brackets to better suit the present discussion):

"So let G be a 3-connected graph, $v_1, v_2, v_3 \in V(G)$ and $n_1 + n_2 + n_3 = |V(G)|$. Take a new point a and connect it to v_1, v_2, and v_3. Consider the topological space $\mathscr{H}$ constructed for this new graph G'. [In our language, $\mathscr{H}$ is the basis complex of the branching greedoid determined by the rooted graph (G', a). This greedoid, whose bases are the spanning trees of G', is 3-connected.] For each spanning tree T of G', let $f_i(T)$ denote the number of points in T accessible from a along the edge $(a, v_i)(i = 1, 2)$. Then the mapping

$$f : T \mapsto (f_1(T), f_2(T))$$

maps the vertices of $\mathscr{H}$ onto lattice points of the plane. Let us subdivide each quadrilateral 2-cell in $\mathscr{H}$ by a diagonal into two triangles; in this way we obtain a triangulation $\bar{\mathscr{H}}$ of $\mathscr{H}$. Extend f affinely to each such triangle so as to obtain a continuous mapping of $\mathscr{H}$ into the plane. Obviously, the image of $\mathscr{H}$ is contained in the triangle $\Delta = \{x \geqslant 0, y \geqslant 0, x + y \leqslant n\}$. We are going to show that the mapping is onto Δ.

"Let us pick three spanning trees, T_1, T_2, T_3 first such that $f(T_1) = (n, 0), f(T_2) = (0, n), f(T_3) = (0, 0)$. Obviously, such trees exist. Next, by applying [the fact that the basis graph of a 2-connected greedoid is connected] to the graph $G' - (a, v_3)$, we select a polygon P_{12} in $\mathscr{H}$ connecting T_1 to T_2 and having $f_3(x) = 0$ at all points. Thus $f(P_{12})$ connects $(n, 0)$ to $(0, n)$ along the side of the triangle Δ with these endpoints. Let P_{23} and P_{31} be defined analogously.

"By Theorem 6.6, $P_{12} + P_{23} + P_{31}$ can be contracted in $\mathscr{H}$ to a single point. Therefore, $f(P_{12}) + f(P_{23}) + f(P_{31})$ can be contracted in $f(\mathscr{H})$ to a single point. But 'obviously' (or, rather, by applying the well-known fact [Brouwer's Theorem 13.1] that the boundary of a triangle cannot be contracted to a single point in the triangle with one interior point taken out), $f(\mathscr{H})$ must cover the whole triangle Δ. So in particular the point (n_1, n_2) belongs to the image of $\mathscr{H}$, and therefore it belongs to the image of a triangle of $\bar{\mathcal{K}}$. But it is easy to see that this implies that (n_1, n_2) is the image of one of the vertices of $\bar{\mathcal{K}}$; i.e., there exists a spanning tree T with

$$f_1(T) = n_1, \quad f_2(T) = n_2.$$

The three components of $T - a$ now yield the desired partition of $V(G)$."

Theorem 6.6 is a special case of a more general result saying that for any k-connected interval greedoid a certain higher-dimensional basis complex is $(k-2)$-connected. This more general result implies Theorem 6.1 for arbitrary k by extension of the ideas we have just seen in the $k = 3$ case. See Lovász (1977) and Björner, Korte and Lovász (1985) for complete details.

Tutte's Homotopy Theorem

A matroid is called *regular* if it can be coordinatized over every field. In Tutte (1958) a characterization is given of regular matroids in terms of forbidden minors. The proof relies in an essential way on a "Homotopy Theorem", expressing the 1-connectivity of certain 2-dimensional complexes. Tutte's Homotopy Theorem was also used by R. Reid and R. Bixby to prove the forbidden minor characterization for representability over GF(3). More recently other proofs of these results, avoiding use of the Homotopy Theorem, have been found by P. Seymour and others. See chapter 10 by Seymour for an up-to-date account.

Tutte's Homotopy Theorem seems to be the oldest topological result of its kind in combinatorics. Unfortunately it is quite technical both to state in full and to prove. Here we shall state the Homotopy Theorem in sufficient detail that the nature of the result can be understood. Complete details can be found in Tutte (1958) and Tutte (1965).

Let L be a finite geometric lattice of rank r, and write L^i for the set of flats of rank i; so L^{r-1} is the set of copoints, L^{r-2} the colines and L^{r-3} the coplanes. Flats $X \in L$ will be thought of as subsets of the point set L^1 via $\bar{X} = \{p \in L^1 \mid p \leqslant X\}$.

Given any point $a \in L^1$ we define a graph $\mathrm{TG}(L, a)$ on the vertex set $L^{r-1}_{\not\geq a} = \{X \in L^{r-1} \mid X \not\geq a\}$ as follows: two copoints X and Y "off a" (i.e., in the set $L^{r-1}_{\not\geq a}$) span an edge if $X \wedge Y$ is a coline and $\bar{X} \cup \bar{Y} \neq L^1 - a$. On this graph we construct a 2-dimensional regular cell complex $\mathrm{TC}(L, a)$ by attaching 2-cells into the triangles and squares of the following kinds:

6.8. Triangles $XYZX$ for which $\mathrm{rk}(X \wedge Y \wedge Z) \geqslant r - 3$.

6.9. Squares $XYZTX$ for which $\mathrm{rk}(P) = r - 3$, where $P = X \wedge Y \wedge Z \wedge T$, and *either* the coline $P \vee a$ is covered by exactly two copoints *or else* the interval $[P, \hat{1}]$ is isomorphic to the lattice of flats of the Fano matroid F_7 minus one of its points.

If L has no minor isomorphic to F_7^*, the dual of the Fano matroid, then (6.8) and (6.9) describe all the 2-cells of the *Tutte complex* $\mathrm{TC}(L, a)$. [This means that for use in representation theory the definition (6.8)–(6.9) of $\mathrm{TC}(L, a)$ is sufficient.] In general it is necessary to attach 2-cells also into certain squares $XYZTX$ for which $\mathrm{rk}(X \wedge Y \wedge Z \wedge T) = r - 4$. The definition of these squares (of the "corank 4 kind") is fairly complicated, so we refrain from describing them here.

Theorem 6.10 (Homotopy Theorem, Tutte 1958). *The complex* $\mathrm{TC}(L,a)$ *is* 1-*connected.*

The combinatorial meaning of Theorem 6.10 is that any two copoints X and Y "off a" can be connected "off a" by a path in the Tutte graph $\mathrm{TG}(L,a)$, and that any two such paths differ by a sequence of *elementary homotopies* of type $XYX, XYZX$ as in (6.8), or $XYZTX$ as in (6.9) or of the corank 4 kind. (Compare the discussion preceding Theorem 6.7.)

The given formulation of the Homotopy Theorem differs in form but not in content from the statement in Tutte (1958). Tutte has remarked about his theorem (Tutte 1979, p. 446) that "the proof ... is long, but it is purely graph-theoretical and geometrical in nature. I am rather surprised that it seems to have acquired a reputation for extreme difficulty". No significant simplification of the original proof seems to be known, other than in special cases. One such case is if $\bar{X} \cup \bar{Y} \neq L^1 - a$ for all pairs X, Y of copoints "off a" such that $X \wedge Y$ is a coline. Then the top three levels of $L - [a, \hat{1}]$ form a poset which is 1-connected by (11.10) (iv), (11.2) and Theorem 11.14, and the 1-connectivity can be transferred to $\mathrm{TC}(L,a)$ by an application of the Fiber Theorem 10.5, similar to the proof of Theorem 6.6. A simpler and more conceptual proof of Tutte's Theorem in full strength would be of definite interest.

Unfortunately the available space does not permit a thorough explanation of how Theorem 6.10 is used in representation theory. Here is a briefest possible sketch of the idea. Tutte's proof of sufficiency for his characterization of regular matroids runs by induction on the size of the ground set (that is why it is of interest to delete the point a). Roughly speaking, the "regular" coordinatization lives on the copoints, and its value at the new point a is extended from one copoint in $L^{r-1}_{\not\geq a}$ to another via paths in the Tutte graph $\mathrm{TG}(L,a)$. The Homotopy Theorem is then needed to check that different paths do not lead to contradictions. A similar idea is illustrated in greater detail in the proof of Theorem 7.6.

7. Oriented matroids

Two topics from the theory of oriented matroids will be discussed in this section. Most important is the topological representation theorem of Folkman and Lawrence (1978), which states that every oriented matroid can be realized by an arrangement of pseudospheres. As an application we show how such realizations lead to quick proofs of some combinatorial properties of rank 3 oriented matroids. Second, we sketch (following Las Vergnas 1978) how Maurer's Homotopy Theorem 6.7 can be used to deduce the existence of a determinantal sign function.

Oriented matroids are defined in chapter 9 by Welsh. Since we will use a slightly different formulation of the concept (due to Folkman and Lawrence 1978) and need to refer to the linear case for motivation, we will start with a quick review of the basics, which will also serve to fix notation. More extensive treatments can be found in the monographs Bachem and Kern (1992) and Björner, Las Vergnas, Sturmfels, White and Ziegler (1993).

Let E be a finite set with a fixed-point free involution $x \mapsto x^*$ (i.e., $x^* \neq x = x^{**}$ for all $x \in E$). Write $A^* = \{x^* \mid x \in A\}$, for subsets $A \subseteq E$. An *oriented matroid* $\mathcal{O} = (E, *, \mathscr{C})$ is such a set together with a family $\mathscr{C}$ of nonempty subsets such that

(OM1) $\mathscr{C}$ is a clutter (i.e., $C_1 \neq C_2$ implies $C_1 \not\subseteq C_2$ for all $C_1, C_2 \in \mathscr{C}$);

(OM2) if $C \in \mathscr{C}$ then $C^* \in \mathscr{C}$ and $C \cap C^* = \emptyset$;

(OM3) if $C_1, C_2 \in \mathscr{C}, C_1 \neq C_2^*$ and $x \in C_1 \cap C_2^*$, then there exists $D \in \mathscr{C}$ such that $D \subseteq C_1 \cup C_2 - \{x, x^*\}$.

The sets in $\mathscr{C}$ are called *circuits* of the oriented matroid $\mathcal{O}$. For elements $x \in E$ let $\bar{x} = \{x, x^*\}$, and let $\bar{A} = \{\bar{x} \mid x \in A\}, A \subseteq E$, and $\bar{\mathscr{C}} = \{\bar{C} \mid C \in \mathscr{C}\}$. The system $\bar{\mathscr{C}}$ satisfies the usual matroid circuit-exchange axioms, so $\bar{\mathcal{O}} = (\bar{E}, \bar{\mathscr{C}})$ is a matroid, called the *underlying matroid* of $\mathcal{O}$. Not all matroids arise from oriented matroids in this way; those that do are called *orientable*. A subset $B \subseteq E$ is called a *basis* of $\mathcal{O}$ if $\bar{B}$ is a basis of $\bar{\mathcal{O}}$. The *rank* of $\mathcal{O}$ equals the rank of $\bar{\mathcal{O}}$. Without significant loss of generality we will make the tacit assumption in what follows that all oriented matroids are *simple*, meaning that no circuit has fewer than three elements.

The fundamental models for oriented matroids are sets of vectors in $\mathbb{R}^d$ and the relation of positive linear dependence (or, more generally, positive linear dependence of vectors over *any* ordered field). Suppose that E is a finite subset of $\mathbb{R}^d - \{0\}$ such that $E = -E$, and if $x \neq y$ in E are parallel then $y = -x$. For $x \in E$ let $x^* = -x$. A subset $A \subseteq E$ is *positive linearly dependent* if $\Sigma_{x \in A} \lambda_x x = 0$ for some real coefficients $\lambda_x \geqslant 0$, not all equal to zero. Let $\mathscr{C}$ be the family of all inclusion-wise minimal positive linearly dependent subsets of E, except those of the form $\{x, x^*\}, x \in E$. Equivalently, $\mathscr{C}$ consists of all subsets of E which form the vertex set of a simplex of dimension $\geqslant 2$ containing the origin in its relative interior. Oriented matroids $(E, *, \mathscr{C})$ which arise in this way are called *linear* (or, *realizable*) over $\mathbb{R}$. Not all oriented matroids are isomorphic to linear ones.

Topological Representation Theorem

To pave the way for the Representation Theorem for oriented matroids it is best to look at the linear case for motivation. The Representation Theorem in fact says that intuition gained from the linear case is going to be essentially correct (modulo some topological deformation which cannot be too bad) for general oriented matroids.

Let E be a finite subset of $\mathbb{R}^d - \{0\}$ such that $E = -E$, and let $\mathcal{O} = (E, *, \mathscr{C})$ be the linear oriented matroid as previously discussed. For each $e \in \bar{E} = \{\bar{x} = \{x, x^*\} \mid x \in E\}$, let H_e be the hyperplane orthogonal to the line spanned by e. The *arrangement of hyperplanes* $\mathscr{H} = \{H_e \mid e \in \bar{E}\}$ contains all information about $\mathcal{O}$, since one can go from H_e back to a pair of opposite normal vectors, and the definition of the sets which form circuits in $\mathcal{O}$ (i.e., the sets in $\mathscr{C}$) is independent of the length of vectors. By intersecting with the unit sphere S^{d-1} we can alternatively look at the *arrangement of spheres* $\mathcal{S} = \{H_e \cap S^{d-1} \mid e \in \bar{E}\}$, which is merely a collection of equatorial $(d-2)$-spheres inside the $(d-1)$-sphere. Clearly: *linear oriented matroids* (up to reorientation), *arrangements of hyperplanes* and *arrangements of spheres* are the same thing.

When thinking about a linear oriented matroid $(E, *, \mathscr{C})$ as an arrangement of spheres it is useful to visualize elements $x \in E$ as closed hemispheres $\bar{H}_x = \{y \in S^{d-1} \mid (y,x) \geqslant 0\}$. Then a subset $A \subseteq E$ belongs to $\mathscr{C}$ if and only if $A \cap A^* = \emptyset$ and A is minimal such that $\bigcup_{x \in A} \bar{H}_x = S^{d-1}$.

We shall need the following terminology. A *sphere* $\sum$ is a topological space for which there is a homeomorphism $f: S^j \to \sum$ with the standard j-sphere $S^j = \{x \in \mathbb{R}^{j+1} \mid \|x\| = 1\}$, for some $j \geqslant 0$. A *pseudosphere* S in $\sum$ is any image $S = f(\{x \in S^j \mid x_{j+1} = 0\})$ under such a homeomorphism. [In the topological literature pseudospheres are known as "tamely embedded (or, flat) codimension-one subspheres", cf. Rushing (1973).] The two *sides* (or, pseudohemispheres) of S are $S^+ = f(\{x \in S^j \mid x_{j+1} \geqslant 0\})$ and $S^- = f(\{x \in S^j \mid x_{j+1} \leqslant 0\})$. Clearly, S is the intersection of its two sides, which are homeomorphic to balls.

The crucial definition is this: An *arrangement of pseudospheres* $(\bar{E}, \mathscr{A})$ in S^{d-1} is a finite collection $\mathscr{A} = \{S_e \mid e \in \bar{E}\}$ of distinct pseudospheres S_e in S^{d-1} such that

(AP1) Every nonempty intersection $S_A = \bigcap_{e \in A} S_e, A \subseteq \bar{E}$, is a sphere.

(AP2) For every nonempty intersection S_A and all $e \in \bar{E}$, either $S_A \subseteq S_e$ or $S_A \cap S_e$ is a pseudosphere in S_A with sides $S_A \cap S_e^+$ and $S_A \cap S_e^-$.

This definition is due to Folkman and Lawrence (1978). They actually required more, but the additional assumptions in their definition were proved to be redundant by Mandel (1982).

In analogy with the linear case (arrangement of spheres), an arrangement of pseudospheres $(\bar{E}, \mathscr{A})$ gives rise to a system $\mathcal{O}(\mathscr{A}) = (E, *, \mathscr{C})$ as follows: put $E = \{S_e^+ \mid e \in \bar{E}\} \cup \{S_e^- \mid e \in \bar{E}\}$, let $(S_e^+)^* = S_e^-$ and vice versa, and define $\mathscr{C}$ to be the collection of the minimal subsets $A \subseteq E$ such that $\bigcup A = S^{d-1}$ and $A \cap A^* = \emptyset$. It turns out that $\mathcal{O}(\mathscr{A})$ is an oriented matroid (in spite of the topological deformations). What is more surprising is that the construction leads to *all* oriented matroids. We call an arrangement $\mathscr{A}$ *essential* if $\bigcap \mathscr{A} = \emptyset$.

Theorem 7.1 (Representation Theorem, Folkman and Lawrence 1978).

(i) *If $\mathscr{A}$ is an arrangement of pseudospheres in S^{d-1}, then $\mathcal{O}(\mathscr{A})$ is an oriented matroid. Furthermore, if $\mathscr{A}$ is essential then rank $\mathcal{O}(\mathscr{A}) = d$.*

(ii) *If $\mathcal{O}$ is an oriented matroid of rank d, then $\mathcal{O} = \mathcal{O}(\mathscr{A})$ for some essential arrangement of pseudospheres in S^{d-1}.*

(iii) *The mapping $\mathscr{A} \to \mathcal{O}(\mathscr{A})$ induces a one-to-one correspondence between rank d oriented matroids and essential arrangements of pseudospheres in S^{d-1}, up to natural equivalence relations.*

The proof of this result is quite involved. For part (ii) a poset is first constructed from the oriented matroid, and then it is shown using Theorem 12.6 that this poset is the poset of faces of some regular cell complex $\mathscr{C}$.This complex $\mathscr{C}$ provides the

$(d-1)$-sphere and various subcomplexes the $(d-2)$-subspheres forming the arrangement. The sphere $\mathscr{C}$ is constructible (Edmonds and Mandel 1978, Mandel 1982), and even shellable (Lawrence 1984), which implies that the whole construction of $\mathscr{C}$ and the relevant subcomplexes can be carried out in piecewise linear topology. In particular, this means that no topological pathologies need to be dealt with in representations of oriented matroids. Complete proofs of Theorem 7.1 can be found in Folkman and Lawrence (1978), Mandel (1982), and Björner, Las Vergnas, Sturmfels, White and Ziegler (1993).

The Representation Theorem shows that oriented matroids of rank 3 correspond to arrangements of "pseudocircles" on the 2-sphere or, in the projective version, arrangements of pseudolines in the real projective plane. This representation can be used for quick proofs of some combinatorial properties as in the following application.

Theorem 7.2. *Let M be an orientable matroid of rank 3. Then:*

(i) *M has a 2-point line,*

(ii) *if the points of M are 2-colored there exists a monochromatic line.*

Here is how Theorem 7.2 follows from Theorem 7.1. Represent the points of M as pseudocircles on the 2-sphere. Then lines are maximal collections of pseudocircles with nonempty intersection (which is necessarily a 0-sphere, i.e., two points). The arrangement of pseudocircles gives a graph G whose vertices are the points of intersection and edges the segments of pseudocircles between such points. Since this graph lies embedded in S^2 it is planar, and since $\mathrm{rk}(M)=3$ it is simple. We need the following lemma.

Lemma 7.3. *For any planarly embedded simple graph:*

(i) *some vertex has degree at most five,*

(ii) *if the edges are 2-colored then there exists a vertex around which the edges of each color class are consecutive in the cyclic ordering induced by the embedding.*

Part (i) is a well-known consequence of Euler's formula (cf. chapter 5 by Thomassen). Part (ii) is also a consequence of Euler's formula, but not as well known. It was used by Cauchy in the proof of his Rigidity Theorem for 3-dimensional convex polytopes.

To finish the proof of Theorem 7.2, look at the graph G determined by the arrangement of pseudocircles. If all lines in M have at least 3 points, then every vertex in G will have degree at least 6, in violation of (i). If the pseudocircles are 2-colored and through every intersection point there is at least one pseudocircle of each color, then the induced coloring of the edges of G will violate (ii).

The proof of the first part of Theorem 7.2, a generalization of the Sylvester–Gallai Theorem (see chapter 17 by Erdős and Purdy), has been known since the 1940s in the linear case. The following strengthening by Csima and Sawyer (1993) also uses pseudoline representation: *The number of 2-point lines in M is at least $\frac{6}{13}(\mathrm{card}\ M)$.* The proof of the second part, due to G.D. Chakerian in the linear case, was rediscovered by Edmonds, Lovász and Mandel (1980), who also observed the generalization to oriented matroids.

Basis signatures

Just like ordinary matroids, oriented matroids can be characterized in several ways. We shall discuss a characteristic property of the set of bases $\mathcal{B}$ of an oriented matroid, namely that a determinant can be defined up to sign (but not magnitude). This was first shown by Las Vergnas (1978). Characterizations of oriented matroids in terms of signed bases were also discovered by J. Bokowski, A. Dress, L. Gutierrez-Novoa and J. Lawrence.

Let us review some essential features of the function $\delta : \tilde{\mathcal{B}} \to \{+1, -1\}$, taking *ordered* bases of a *linear* oriented matroid $(E, *, \mathcal{C}), E \subseteq \mathbb{R}^d$, to the sign of their determinants. A function η can be defined for certain pairs of ordered bases β and β' in $\mathbb{R}^d$ as follows:

7.4. Suppose β and β' are permutations of the same basis B. Let $\eta(\beta, \beta') = +1$ if they are of the same parity and $= -1$ otherwise.

7.5. Suppose $\beta = x_1 x_2 \cdots x_{r-1} y$ and $\beta' = x_1 x_2 \cdots x_{r-1} z$ with $y \neq z$. Let $\eta(\beta, \beta') = +1$ if y and z are on the same side of the hyperplane spanned by $\{x_1, \ldots, x_{r-1}\}$, and $= -1$ otherwise.

Now, once we choose an ordered basis β_0 and put $\det(\beta_0) := +1$, the function $\det(\beta)$ and its sign $\delta(\beta)$ is determined for all ordered bases β by the usual rules of linear algebra. But the function $\delta(\beta)$ is also *combinatorially* determined, because any pair of ordered bases can be connected by a chain of steps of type (7.4) or (7.5) and we have: *If β and β' are ordered bases as in* (7.4) *or* (7.5) *then* $\delta(\beta) = \eta(\beta, \beta') \cdot \delta(\beta')$.

The preceding discussion points the way how to generalize the determinantal sign function to all oriented matroids. First, to cast (7.5) in a form which is more compatible with the axiom system (OM 1)–(OM 3), we replace it by the following reformulation:

7.5′. Suppose $\beta = x_1 x_2 \cdots x_{r-1} y$ and $\beta' = x_1 x_2 \cdots x_{r-1} z$ with $y \neq z$, and if $y \neq z^*$ let $\{C, C^*\}$ be the unique pair of circuits such that in the underlying matroid $\{\bar{y}, \bar{z}\} \subseteq \bar{C} \subseteq \{\bar{x}_1, \ldots, \bar{x}_r, \bar{y}, \bar{z}\}$. Put $\eta(\beta, \beta') = +1$ if one of y and z lies in C and the other in C^*, and put $\eta(\beta, \beta') = -1$ otherwise.

Theorem 7.6 (Las Vergnas 1978). *Let $\tilde{\mathcal{B}}$ be the set of ordered bases of an oriented matroid, and let $\beta_0 \in \tilde{\mathcal{B}}$. There exists a unique function $\delta : \tilde{\mathcal{B}} \to \{+1, -1\}$ such that $\delta(\beta_0) = +1$ and if $\beta, \beta' \in \tilde{\mathcal{B}}$ are related as in* (7.4) *or* (7.5′) *then* $\delta(\beta) = \eta(\beta, \beta') \cdot \delta(\beta')$.

The proof runs as follows. Define a graph on the vertex set $\tilde{\mathcal{B}}$ by connecting pairs $\{\beta, \beta'\}$ which are related as in (7.4) or (7.5′) by an edge. The graph is clearly connected, and there is a projection $\pi : \tilde{\mathscr{B}} \to \mathscr{B}$ to the basis graph $\mathscr{B}$ of the underlying matroid. Now, put $\delta(\beta_0) := +1$, and for $\beta \in \tilde{\mathscr{B}}$ define

$$\delta(\beta) := \prod_{i=1}^{n} \eta(\beta_{i-1}, \beta_i)$$

for some choice of path $\beta_0, \beta_1, \ldots, \beta_n = \beta$ in $\tilde{\mathscr{B}}$. The proof is complete once we show that this definition is independent of the choice of path from β_0 to β. If P_1 and P_2 are two such paths then by Theorem 6.7 their projections $\pi(P_1)$ and $\pi(P_2)$ in the basis graph differ by a sequence of elementary homotopies. Thus the checking is reduced to verifying

$$\prod_{i=1}^{k} \eta(\alpha_{i-1}, \alpha_i) = 1$$

for closed paths $\alpha_0, \alpha_1, \ldots, \alpha_k = \alpha_0$ in $\tilde{\mathscr{B}}$ whose projection in $\mathscr{B}$ is an edge BCB, triangle $BCDB$ or square $BCDEB$. However, the basis configurations which give triangles or squares in the basis graph are explicitly characterized in (6.3)–(6.5), and this way the checking is brought down to a manageable size. See Las Vergnas (1978) for further details.

8. Discrete applications of the Hard Lefschetz Theorem

One of the most esoteric results to have found applications in combinatorics is the Hard Lefschetz Theorem. It was used by R. Stanley to prove the Erdős–Moser conjecture (chapter 32 by Alon) and to show necessity in the characterization of f-vectors of simplicial convex polytopes (chapter 18 by Klee and Kleinschmidt).

In this section we will state the Hard Lefschetz Theorem and briefly explain how it is used for these applications. The presentation follows Stanley (1980a,b, 1983b, 1985, 1989). Other applications appear in Stanley (1987a,b).

Unfortunately, concepts must be used here which go beyond what is reviewed and explained in part II of this chapter. In particular we must assume some familiarity with the singular cohomology ring of a topological space, and with a few basic notions of algebraic geometry (projective varieties, smoothness, etc.). See Hartshorne (1977) for this.

Let X be a smooth irreducible complex projective variety of complex dimension d, and let $H^*(X) = H^0(X) \oplus H^1(X) \oplus \cdots \oplus H^{2d}(X)$ denote its singular cohomology ring with real coefficients. Recall that if $\omega \in H^i(X)$ and $\tau \in H^j(X)$ then $\omega \cdot \tau \in H^{i+j}(X)$. Being projective, we may intersect X with a generic hyperplane H of an ambient projective space. By a standard construction in algebraic geometry the subvariety $X \cap H$ represents a cohomology class $\omega \in H^2(X)$.

Theorem 8.1 (The Hard Lefschetz Theorem). *Let X and $\omega \in H^2(X)$ be as above, and let $0 \leqslant i \leqslant d$. Then the linear map $H^i(X) \to H^{2d-i}(X)$ given by multiplication by ω^{d-i} is an isomorphism of vector spaces.*

See Stanley (1983b) for references to various proofs of this theorem (the first rigorous one is due to W. Hodge). Note that the fact that $H^i(X)$ and $H^{2d-i}(X)$ are isomorphic is known already from Poincaré duality. Thus the point of the theorem is entirely the existence of a special cohomology class ω with such favorable multiplicative properties. Whereas Poincaré duality is a purely topological phenomenon (valid for all compact orientable manifolds, and in various versions also more generally), the Hard Lefschetz Theorem uses smoothness in an essential way. There is not (as far as is known) any intrinsically topological construction of a good cohomology class ω that would make Theorem 8.1 valid for some reasonable class of topological manifolds. Nevertheless, the Hard Lefschetz Theorem has been extended to some more general classes of varieties, e.g., to Kähler manifolds in differential topology and to V-varieties (nonsmooth varieties with finite quotient singularities, e.g,, the toric varieties of simplicial polytopes discussed below).

Stanley's (1980a) proof of the Erdős–Moser conjecture is outlined in section 9 of chapter 32 by Alon. Referring to the discussion there, and using the same notation, we will now indicate how Theorem 8.1 is used.

For a certain poset $M(n)$ of rank $N = \binom{n+1}{2}$ and with rank-level sets $M(n)_i, i = 0, 1, \ldots, N$, let V_i be the real vector space with basis $M(n)_i$. For the proof it is needed to construct linear mappings $\varphi_i : V_i \to V_{i+1}$ such that the composition $\varphi_{N-i-1} \circ \varphi_{N-i-2} \circ \cdots \circ \varphi_i : V_i \to V_{N-i}$ is invertible, for $0 \leqslant i \leqslant [N/2]$, and if $x \in M(n)_i$ and $\varphi_i(x) = \sum_{y \in M(n)_{i+1}} c_y \cdot y$, then $c_y \neq 0$ implies $y > x$.

Take the special orthogonal group $G = \mathrm{SO}_{2n+1}(\mathbb{C})$ and let P be the maximal parabolic subgroup corresponding to the simply-laced part of its Dynkin diagram. Then G/P is a smooth irreducible complex projective variety having a cell decomposition (in a certain algebraic-geometric sense) such that the poset of closed cells is isomorphic to $M(n)$. This cell decomposition of G/P (induced by the Bruhat decomposition of G) has cells only in even dimensions, and we may identify $M(n)_i$ with the set of $2i$-dimensional cells and conclude that $V_i \cong H^{2i}(G/P)$. The relevance of Theorem 8.1 is now becoming clear; indeed, letting the linear mapping $\varphi_i : V_i \to V_{i+1}$ be multiplication with ω, all required properties turn out to hold.

The poset $M(n)$ is a member of a class of finite rank-symmetric posets arising as Bruhat order on Weyl groups and on their quotients modulo parabolic subgroups. Using Theorem 8.1, Stanley (1980a) showed that all such posets are rank-unimodal and satisfy a strong form of the Sperner property.

Many of the results of Stanley (1980a), including the proof of the Erdős–Moser conjecture, can be proven with just linear algebra, see Proctor (1982). This is done, essentially, by rewriting the first proof (including a proof of the Hard Lefschetz Theorem) as concretely as possible and throwing out all mention of algebraic geometry.

We now turn to the characterization of f-vectors of simplicial polytopes. This application of Theorem 8.1 uses more of its content. The fact that the linear mappings φ_i constructed above are given by multiplication is irrelevant for the previous argument, whereas the global multiplicative structure of $H^{\bullet}(X)$ is essential in what follows.

We refer to chapter 18 by Klee and Kleinschmidt for definitions relating to simplicial d-polytopes P and their h-vectors $h(P) = (h_0, h_1, \ldots, h_d)$. As observed there, every simplicial polytope in $\mathbb{R}^d$ is combinatorially equivalent to one with vertices in $\mathbb{Q}^d$.

Let P be a d-dimensional convex polytope with vertices in $\mathbb{Q}^d$. There is a general construction (see Ewald 1995, Fulton 1993 or Oda 1988) which associates with P an irreducible complex projective variety $X(P)$ of complex dimension d, called a *toric variety*. This variety is in general not smooth, not even in the simplicial case.

Suppose now that P is simplicial. Then the following is true [work of V.I. Danilov, J. Jurkiewicz, M. Saito and others; see the cited books or Stanley (1983b, 1985, 1987a)]:

(i) the cohomology of $X(P)$ vanishes in all odd dimensions, and $\dim_{\mathbb{R}} H^{2i}(X(P)) = h_i(P)$, for $i = 0, 1, \ldots, d$.

(ii) $H^*(X(P))$ is generated (as an algebra over $\mathbb{R}$) by $H^2(X(P))$,

(iii) the Hard Lefschetz Theorem 8.1 holds for $X{=}X(P)$ and the class of a hyperplane section $\omega \in H^2(X)$.

It follows from (iii) that the mapping $H^{2i}(X) \to H^{2(i+1)}(X)$ given by multiplication with ω is injective if $i < d/2$ and surjective if $i \geqslant [d/2]$. Therefore, taking the quotient of the cohomology ring

$$H^*(X) = \oplus_{i=0}^{d} H^{2i}(X)$$

by the ideal generated by ω, we get a graded ring

$$R = H^*(X)/\langle\omega\rangle = \oplus_{i=0}^{[d/2]} R_i,$$

where $R_i = H^{2i}(X)/\omega H^{2i-2}(X)$, for $i \geqslant 1$, and $R_0 = H^0(X) \cong \mathbb{R}$. Furthermore, R is generated by R_1 [by (ii)], and $\dim_{\mathbb{R}} R_i = h_i - h_{i-1}$ [by (i) and (iii)]. This shows that $(h_0, h_1 - h_0, h_2 - h_1, \ldots, h_{[d/2]} - h_{[d/2]-1})$ is an "$\mathcal{O}$-sequence", as defined in Theorem 6.2 of chapter 18 by Klee and Kleinschmidt. As explained after Theorem 6.5 of that chapter, this is precisely what needs to be shown to complete the proof of necessity of the characterization of f-vectors of simplicial polytopes.

A more elementary (and self-contained) proof of necessity has recently been found by McMullen (1993). He replaces the cohomology ring of the toric variety by a certain subalgebra of the polytope algebra and proves the needed analog of the Hard Lefschetz Theorem using convex geometry.

In Stanley (1987a) sharp lower bounds are given for the differences $h_i - h_{i-1}$, $1 \leqslant i \leqslant [d/2]$, for a centrally symmetric simplicial d-polytope. The proof involves the interaction between the Hard Lefschetz Theorem and a finite group action.

The toric variety $X{=}X(P)$ of a *non-simplicial* polytope P with rational vertices is unfortunately more difficult to use for combinatorial purposes. For instance, $\dim_{\mathbb{R}} H^i(X)$ may depend on the embedding of P and not only on its combinatorial type, and cohomology may fail to vanish in odd dimensions. However, the *intersection cohomology* (of middle perversity) $IH^*(X)$, defined by M. Goresky and R. MacPherson, turns out to be combinatorial and to satisfy a module version of the hard Lefschetz theorem. This leads to some interesting information for general rational polytopes, such as Theorem 6.8 of chapter 18 by Klee and Kleinschmidt. See Stanley (1987b) for more information.

PART II. TOOLS

The rest of this chapter is devoted to a review of some definitions and results from combinatorial topology that have proven to be particularly useful in combinatorics. The material in sections 9 (simplicial complexes), 12 (cell complexes) and 13 (fixed-point and antipodality theorems) is of a very general nature and detailed treatments can be found in many topology books. Specific references will therefore be given only sporadically. Most topics in sections 10 and 11, on the other hand, are of a more special nature, and more substantial references (and even some proofs) will be given.

Many of the results mentioned have been discussed in a large number of papers and books. When relevant, our policy has been to reference the original source (when known to us) and some more recent papers that contribute simple proofs, extensions or up-to-date discussion (a subjective choice). We apologize for any inaccuracy or omission that may unintentionally have occurred.

9. Combinatorial topology

This section will review basic facts concerning simplicial complexes. Good general references are Munkres (1984a) and Spanier (1966). Basic notions such as *(topological) space*, *continuous map* and *homeomorphism* will be considered known. Throughout this chapter, every map between topological spaces is assumed to be continuous, even if not explicitly stated.

Simplicial complexes and posets

9.1. An *(abstract) simplicial complex* $\Delta = (V, \Delta)$ is a set V (the *vertex set*) together with a family Δ of nonempty finite subsets of V (called *simplices* or *faces*) such that $\emptyset \neq \sigma \subseteq \tau \in \Delta$ implies $\sigma \in \Delta$. Usually, $V = \bigcup \Delta$ (shorthand for $V = \bigcup_{\sigma \in \Delta} \sigma$) so V can be suppressed from the notation.

The *dimension* of a face σ is $\dim \sigma = \operatorname{card} \sigma - 1$, the *dimension* of Δ is $\dim \Delta = \max_{\sigma \in \Delta} \dim \sigma$. A d-dimensional complex is *pure* if every face is contained in a *d-face* (i.e., d-dimensional face). The complex consisting of all nonempty subsets of a $(d+1)$-element set is called the *d-simplex*.

Note that our definition allows the empty complex $\Delta = \emptyset$. It is, by convention, (-1)-dimensional. [Remark: The definition of a simplicial complex (with *nonempty* faces) that we use here is the standard one in topology. In combinatorics it is usually more convenient to allow the empty set as a face of a complex; in particular, this is consistent with the definition of reduced homology.]

Let $\Delta^k = \{k\text{-faces of } \Delta\}$ and $\Delta^{\leqslant k} = \bigcup_{j \leqslant k} \Delta^j$, for $k \geqslant 0$. The elements of $\Delta^0 = V$ and Δ^1 are called *vertices* and *edges*, respectively. If Δ is pure d-dimensional the elements of Δ^d are called *facets* (or *chambers*). $\Delta^{\leqslant k}$ is the *k-skeleton* of Δ. It is a *subcomplex* of Δ.

A (*geometric*) *simplicial complex* is a polyhedral complex in $\mathbb{R}^d$ [in the sense of (12.1)] whose cells are geometric simplices (the convex hull of affinely independent point-sets). If Γ is a geometric simplicial complex then the family of extreme-point-sets of cells in Γ form an abstract simplicial complex $\Delta(\Gamma)$ which is finite. Conversely, if $\Delta \neq \emptyset$ is a d-dimensional finite abstract simplicial complex then there exist geometric simplicial complexes Γ in $\mathbb{R}^{2d+1}$ such that $\Delta(\Gamma) \cong \Delta$. The underlying space $\bigcup \Gamma$ of any such Γ, unique up to linear homeomorphism, is called the *geometric realization* (or *space*) of Δ, denoted by $\|\Delta\|$. Conversely, Δ is called a *triangulation* of the space $\|\Delta\|$, and of every space homeomorphic to it. Thus, abstract and geometric simplicial complexes are equivalent notions in the finite case (and more generally, when finite-dimensional, denumerable and locally finite). The geometric realization $\|\Delta\|$ of arbitrary infinite abstract simplicial complexes Δ can be constructed as in Spanier (1966).

A *simplicial map* $f: \Delta_1 \to \Delta_2$ is a mapping $f: \Delta_1^0 \to \Delta_2^0$ such that $f(\sigma) \in \Delta_2$ for all $\sigma \in \Delta_1$. By affine extension across simplices it induces a continuous map $\|f\|: \|\Delta_1\| \to \|\Delta_2\|$.

Whereas the rectilinear realization of all d-dimensional simplicial complexes in $\mathbb{R}^{2d+1}$ is easy to prove (and $2d+1$ is best possible), the existence in special cases of rectilinear and of topological realizations in spaces $\mathbb{R}^j$, for $d < j \leqslant 2d$, are difficult and much studied problems. For $d = 1$ this is the question of planarity of graphs (see chapter 5 by Thomassen), for rectilinear embeddings when $d \geqslant 2$, see, e.g., Bokowski and Sturmfels (1989) and the references found therein, and for topological embeddings see Rushing (1973). It is for instance not known whether every triangulation of the 2-dimensional torus has a rectilinear embedding into $\mathbb{R}^3$. A classical result concerning topological embeddings is the van Kampen–Flores Theorem (from 1932–33), which says that the d-skeleton of a $(2d+2)$-simplex does not embed into $\mathbb{R}^{2d}$. Sarkaria (1991b) gives an up-to-date discussion of this result in a setting which also includes the topological Radon–Tverberg theorems discussed in section 5, see also Sarkaria (1991a).

9.2. Let $P = (P, \leqslant)$ be a *poset* (partially ordered set). A totally ordered subset $x_0 < x_1 < \cdots < x_k$ is called a *chain* of *length* k. The supremum of this number over all chains in P is the *rank* (or *length*) of P. If all maximal chains have the same finite length then P is *pure*. P is a *lattice* if every pair of elements $x, y \in P$ has a least upper bound (*join*) $x \vee y$ and a greatest lower bound (*meet*) $x \wedge y$.

For $x \in P$, let $P_{\geqslant x}, P_{>x}, P_{\leqslant x}, P_{<x}$ be defined by $P_{\geqslant x} = \{y \in P : y \geqslant x\}$, etc. For $x \leqslant y$ define the *open interval* $(x,y) = P_{>x} \cap P_{<y}$ and the *closed interval* $[x,y] = P_{\geqslant x} \cap P_{\leqslant y}$. A *bottom element* $\hat{0}$ and a *top element* $\hat{1}$ in P are elements satisfying $\hat{0} \leqslant x$ (respectively $x \leqslant \hat{1}$) for all $x \in P$. If both $\hat{0}$ and $\hat{1}$ exist, P is *bounded*. Then

$\bar{P} = P - \{\hat{0}, \hat{1}\}$ denotes the *proper part* of P. For arbitrary poset $P, \hat{P} = Pcup\{\hat{0}, \hat{1}\}$ denotes P extended by new top and bottom elements (so, card $(\hat{P} \backslash P) = 2$).

Let P be a pure poset of rank r. For $x \in P$, let $r(x) =$ rank$(P_{\leqslant x})$. The *rank function* $r : P \to \{0, 1, \ldots, r\}$ is bijective on each maximal chain. It decomposes P into *rank levels* $P^i = \{x \in P : r(x) = i\}, 0 \leqslant i \leqslant r$.

9.3. The *face poset* $P(\Delta) = (\Delta, \subseteq)$ of a simplicial complex Δ is the set of faces ordered by inclusion. The *face lattice* of Δ is $\hat{P}(\Delta) = P(\Delta) \cup \{\hat{0}, \hat{1}\}$. It is a lattice. $P(\Delta)$ is pure iff Δ is pure, and rank $P(\Delta) = \dim \Delta$.

The *order complex* $\Delta(P)$ of a poset P is the simplicial complex on vertex set P whose k-faces are the k-chains $x_0 < x_1 < \cdots < x_k$ in P. A poset map $f : P_1 \to P_2$ which is *order-preserving* [$x \leqslant y$ implies $f(x) \leqslant f(y)$] or *order-reversing* [$x \leqslant y$ implies $f(x) \geqslant f(y)$] is simplicial $f : \Delta(P_1) \to \Delta(P_2)$, and therefore induces a continuous map $\|f\| : \|\Delta(P_1)\| \to \|\Delta(P_2)\|$. The definition of $\Delta(P)$ goes back to Aleksandrov (1937).

For a simplicial complex Δ, sd$\Delta = \Delta(P(\Delta))$, is called the *(first) barycentric subdivision* (due to its geometric version). A basic fact is that Δ and sdΔ are homeomorphic. Therefore, passage between simplicial complexes and posets via the mappings $P(\cdot)$ and $\Delta(\cdot)$ does not affect the topology, and from a topological point of view simplicial complexes and posets can be considered to be essentially equivalent notions.

The geometric realization $\|P\| = \|\Delta(P)\|$ associates a topological space with every poset P. In this chapter, whenever we make topological statements about a poset P we have the space $\|P\|$ in mind.

There exists at least one other way of associating a useful topology with a poset P (also due to Aleksandrov 1937), namely, let the *order-ideals* (subsets $A \subseteq P$ satisfying $x \leqslant y \in A$ implies $x \in A$) be the open sets of a topology on P. Denote this space $T(P)$. For instance, for the poset depicted to the right in fig. 2 (section 12), $T(P)$ is a space with exactly ten open sets, whereas $\Delta(P)$ is homeomorphic to the 2-sphere. For the *ideal topology* $T(\cdot)$ the continuous maps are precisely the order-preserving maps and homotopy [see (9.10)] has a direct combinatorial meaning. For instance, $T(P)$ is contractible iff P is dismantlable in the sense of (11.1); see Stong (1966). The ideal topology $T(P)$ is relevant for sheaf cohomology over posets (Baclawski 1975, Yuzvinsky 1987) and has surprising connections with the order complex topology $\Delta(P)$ (McCord 1966).

9.4. Let T be a topological space, $\approx$ an equivalence relation on T, and $\pi : T \to T/\!\approx$ the projection map. The quotient $T/\!\approx$ is made into a topological space by letting $A \subseteq T/\!\approx$ be open iff $\pi^{-1}(A)$ is open in T. If $S_i, i \in I$, are pairwise disjoint subsets of T, then $T/(S_i)_{i \in I}$ denotes the *quotient space* obtained by identifying the points within each set $S_i, i \in I$. For example, cone$(T) = T \times [0,1]/(T \times \{1\})$ is the *cone over* T, and susp$(T) = T \times [0,1]/(T \times \{0\}, T \times \{1\})$ is the *suspension* of T. The d-ball modulo its boundary is homeomorphic to the d-sphere: $\boldsymbol{B}^d/\boldsymbol{S}^{d-1} \cong \boldsymbol{S}^d$.

If $(T_i, x_i)_{i\in I}, x_i \in T_i$, is a family of pointed pairwise-disjoint spaces, then the *wedge* of this family is $\bigcup_{i\in I} T_i/(\bigcup_{i\in I}\{x_i\})$. The *join* of two spaces T_1 and T_2 is the space $T_1 * T_2 = T_1 \times T_2 \times [0,1]/(\{(t,x,0) \mid x \in T_2\}, \{(y,s,1) \mid y \in T_1\})_{t\in T_1, s\in T_2}$.

The *join* of two simplicial complexes Δ_1 and Δ_2 (with $\Delta_1^0 \cap \Delta_2^0 = \emptyset$) is the complex $\Delta_1 * \Delta_2 = \Delta_1 \cup \Delta_2 \cup \{\sigma \cup \tau \mid \sigma \in \Delta_1 \text{ and } \tau \in \Delta_2\}$. Further, the *cone* over Δ and *suspension* of Δ are the complexes $\text{cone}(\Delta) = \Delta * \Gamma_1$, $\text{susp}(\Delta) = \Delta * \Gamma_2$, where Γ_i is the 0-dimensional complex with i vertices, $i = 1, 2$. There is a homeomorphism

$$\|\Delta_1 * \Delta_2\| \cong \|\Delta_1\| * \|\Delta_2\|. \tag{9.5}$$

[In case Δ_1 and Δ_2 are not locally finite the topology of the right-hand side may need to be modified to the associated compactly generated topology, see Walker (1988).] In particular, $\|\text{cone}(\Delta)\| \cong \text{cone}(\|\Delta\|)$ and $\|\text{susp}(\Delta)\| \cong \text{susp}(\|\Delta\|)$.

The join of two complexes Δ_1 and Δ_2 has the following geometric realization. First realize Δ_1 and Δ_2 in the same space $\mathbb{R}^d$, with d sufficiently large, so that two distinct line segments $[x_1, x_2]$ and $[y_1, y_2]$ with $x_1, y_1 \in \|\Delta_1\|$ and $x_2, y_2 \in \|\Delta_2\|$ never intersect in an interior point. Then take the union of all such line segments (with the topology induced as a subspace of $\mathbb{R}^d$) – this gives $\|\Delta_1 * \Delta_2\|$.

The *p-fold deleted join* $\Delta_*^{(p)}$ of a simplicial complex Δ is defined as follows. Let $\Delta_1, \ldots, \Delta_p$ be disjoint copies of Δ with isomorphisms $f_i : \Delta_i \to \Delta$. Then $\Delta_*^{(p)}$ is the subcomplex of $\Delta_1 * \cdots * \Delta_p$ consisting of all faces $\sigma_1 \cup \cdots \cup \sigma_p$ such that $f_i(\sigma_i) \cap f_j(\sigma_j) = \emptyset$ for all $i \neq j$. For combinatorial uses of this construction see Sarkaria (1990, 1991a,b) and Živaljević and Vrećica (1992).

The *direct product* $P \times Q$ of two posets is the Cartesian product set ordered by $(x,y) \leqslant (x',y')$ if $x \leqslant x'$ in P and $y \leqslant y'$ in Q. The *join* (or *ordinal sum*) $P * Q$ of two posets is their disjoint union ordered by making each element of P less than each element of Q and otherwise keeping the given orderings within P and Q. Clearly, $\Delta(P * Q) = \Delta(P) * \Delta(Q)$.

There are the following homeomorphisms (Quillen 1978, Walker 1988):

$$\|P \times Q\| \cong \|P\| \times \|Q\|, \tag{9.6}$$

$$\|(P \times Q)_{>(x,y)}\| \cong \|P_{>x}\| * \|Q_{>y}\|, \tag{9.7}$$

$$\|((x,y),(x',y'))\| \cong \text{susp}(\|(x,x')\| * \|(y,y')\|),$$
$$\text{if } x < x' \text{ in } P \text{ and } y < y' \text{ in } Q. \tag{9.8}$$

(Again, special care has to be taken with the topology of the right-hand sides if the participating order complexes are not locally finite.)

9.9. Let Δ be a simplicial complex and $\sigma \in \Delta \cup \{\emptyset\}$. Then define the subcomplexes: *deletion* $\text{dl}_\Delta(\sigma) = \{\tau \in \Delta \mid \tau \cap \sigma = \emptyset\}$, *star* $\text{st}_\Delta(\sigma) = \{\tau \in \Delta \mid \tau \cup \sigma \in \Delta\}$ and *link* $\text{lk}_\Delta(\sigma) = \{\tau \in \Delta \mid \tau \cap \sigma = \emptyset \text{ and } \tau \cup \sigma \in \Delta\}$. Clearly, $\text{dl}(\sigma) \cap \text{st}(\sigma) = \text{lk}(\sigma)$ and $\sigma * \text{lk}(\sigma) = \text{st}(\sigma)$. If $\sigma \in \Delta^0$ then also $\text{dl}(\sigma) \cup \text{st}(\sigma) = \Delta$; and $\text{dl}(\emptyset) = \text{st}(\emptyset) = \text{lk}(\emptyset) = \Delta$.

Homotopy and homology

9.10. Two mappings $f_0, f_1 : T_1 \to T_2$ of topological spaces are *homotopic* (written $f_0 \sim f_1$) if there exists a mapping (called a *homotopy*) $F : T_1 \times [0,1] \to T_2$ such that $F(t,0) = f_0(t)$ and $F(t,1) = f_1(t)$ for all $t \in T_1$. (Remember that all mappings between topological spaces are assumed to be continuous.) The spaces T_1 and T_2 are *of the same homotopy type* (or *are homotopy equivalent*) if there exist mappings $f_1 : T_1 \to T_2$ and $f_2 : T_2 \to T_1$ such that $f_2 \circ f_1 \sim \mathrm{id}_{T_1}$ and $f_1 \circ f_2 \sim \mathrm{id}_{T_2}$. Denote this by $T_1 \simeq T_2$. A space which is homotopy equivalent to a point is called *contractible*.

Let $\boldsymbol{S}^{d-1} = \{x \in \mathbb{R}^d \mid \|x\| = 1\}$ and $\boldsymbol{B}^d = \{x \in \mathbb{R}^d \mid \|x\| \leqslant 1\}$ denote the standard $(d-1)$-*sphere* and d-*ball*, respectively. Note that $\boldsymbol{S}^{-1} = \emptyset$, $\boldsymbol{S}^0 = \{\text{two points}\}$ and $\boldsymbol{B}^0 = \{\text{point}\}$. The class of spheres and balls is closed under the operation of taking joins (up to homeomorphism): $\boldsymbol{S}^a * \boldsymbol{S}^b \cong \boldsymbol{S}^{a+b+1}$, $\boldsymbol{B}^a * \boldsymbol{B}^b \cong \boldsymbol{B}^a * \boldsymbol{S}^b \cong \boldsymbol{B}^{a+b+1}$.

A space T is k-*connected* if for all $0 \leqslant i \leqslant k$ each mapping $f : \boldsymbol{S}^i \to T$ can be extended to a mapping $\hat{f} : \boldsymbol{B}^{i+1} \to T$ such that $\hat{f}(x) = f(x)$ for all $x \in \boldsymbol{S}^i$. In particular, 0-connected means arcwise connected. The property of being k-connected is a *homotopy invariant* (i.e., is transferred to other spaces of the same homotopy type). $\boldsymbol{S}^d$ is $(d-1)$-connected but not d-connected (see Theorem 13.1), $\boldsymbol{B}^d$ is contractible. It is convenient to define the following degenerate cases: (-1)-*connected* means "nonempty", and every space (whether empty or not) is k-*connected* for $k \leqslant -2$.

A simplicial complex Δ is contractible iff Δ is k-connected for all $k \geqslant 0$ (or equivalently, for all $0 \leqslant k \leqslant \dim \Delta$). (The corresponding statement for general spaces is false in the nontrivial direction.) Furthermore, a simplicial complex is k-connected iff its $(k+1)$-skeleton is k-connected.

Let $\pi_i(T) = \pi_i(T,x)$ denote the set of homotopy classes of maps $f : \boldsymbol{S}^i \to T$ such that $f((1,0,\ldots,0)) = x$, from the pointed i-sphere to a pointed topological space $(T,x), x \in T$, $i \geqslant 0$. For $i \geqslant 1$ there exists a composition that makes $\pi_i(T)$ into a group, the ith *homotopy group* of T (at the point x). For $i \geqslant 2$, the group $\pi_i(T)$ is Abelian. $\pi_1(T)$ is the *fundamental group*, and T is *simply connected* if $\pi_1(T) = 0$. The space T is k-connected iff $\pi_i(T,x) = 0$ for all $0 \leqslant i \leqslant k$ and $x \in T$. So, 1-connected means simply connected and arcwise connected.

9.11. For the definitions of *simplicial homology groups* $H_i(\Delta, G)$ and *reduced simplicial homology groups* $\tilde{H}_i(\Delta, G)$ of a complex Δ with coefficients in an Abelian group G, we refer to Munkres (1984a) or Spanier (1966).

Let $\tilde{H}_i(\Delta) = \tilde{H}_i(\Delta, \mathbb{Z})$. The degenerate case

$$\tilde{H}_i(\emptyset) \cong \begin{cases} \mathbb{Z}, & i = -1, \\ 0, & i \neq -1, \end{cases}$$

should be noted. For $\Delta \neq \emptyset$, $\tilde{H}_i(\Delta) = 0$ for all $i < 0$ and all $i > \dim \Delta$, and $\tilde{H}_0(\Delta) \cong \mathbb{Z}^{c-1}$, where c is the number of connected components of Δ. $H_i(\Delta) = \tilde{H}_i(\Delta)$ for all $i \neq -1, 0$; $H_{-1}(\Delta) = 0$ and $H_0(\Delta) \cong \tilde{H}_0(\Delta) \oplus \mathbb{Z}$.

Let Δ_1 and Δ_2 be finite complexes and assume that at least one of $\tilde{H}_p(\Delta_1)$ and $\tilde{H}_q(\Delta_2)$ is torsion-free when $p + q = i - 1$. Then

$$\tilde{H}_{i+1}(\Delta_1 * \Delta_2) \cong \bigoplus_{p+q=i} (\tilde{H}_p(\Delta_1) \otimes \tilde{H}_q(\Delta_2)). \tag{9.12}$$

The same decomposition holds (without any restriction) for reduced homology with coefficients in a field. See Milnor (1956) or chapter V of Cooke and Finney (1967) for further details.

For a finite simplicial complex Δ let $\beta_i = \operatorname{rank} H_i(\Delta) = \dim_{\mathbb{Q}} H_i(\Delta, \mathbb{Q}), i \geqslant 0$. The *Betti numbers* β_i satisfy the *Euler–Poincaré formula*

$$\sum_{i \geqslant 0} (-1)^i \operatorname{card}(\Delta^i) = \sum_{i \geqslant 0} (-1)^i \beta_i. \tag{9.13}$$

Either side of (9.13) can be taken as the definition of the *Euler characteristic* $\chi(\Delta)$. The *reduced Euler characteristic* is $\tilde{\chi}(\Delta) = \chi(\Delta) - 1$. Formula (9.13) is valid with $\beta_i = \dim_{\boldsymbol{k}} H_i(\Delta, \boldsymbol{k})$ for an arbitrary field $\boldsymbol{k}$, although the individual integers β_i may depend on $\boldsymbol{k}$. Additional relations exist between the face-count numbers $f_i = \operatorname{card}(\Delta^i)$ and the Betti numbers β_i (Björner and Kalai 1988). Much is known about the *f-vectors* $f(\Delta) = (f_0, f_1, \ldots)$ for various special classes of complexes Δ. See chapter 18 by Klee and Kleinschmidt for the important case of polytope boundaries, and Björner and Kalai (1989) for a survey devoted to more general classes of complexes.

The Möbius function of a (locally) finite poset is defined in chapter 21 by Gessel and Stanley. Theorem 13.4 of that chapter (due to P. Hall) can in view of (9.13) be restated as

$$\mu(x,y) = \tilde{\chi}(\Delta((x,y))), \quad \text{if } x < y, \tag{9.14}$$

where the right-hand side denotes the reduced Euler characteristic of the order complex of the open interval (x,y). This connection between the Möbius function and topology, first pointed out by Rota (1964) and Folkman (1966), has many interesting ramifications.

9.15. Two complexes of the same homotopy type have isomorphic homology groups in all dimensions. A complex Δ is *k-acyclic over* $\boldsymbol{G}$ if $\tilde{H}_i(\Delta, \boldsymbol{G}) = 0$ for all $i \leqslant k$. So, (-1)-*acyclic* means nonempty and 0-*acyclic* means nonempty and connected. Further, Δ is *acyclic* over $\boldsymbol{G}$ (or simply "$\boldsymbol{G}$-acyclic" if confusion cannot arise) if $\tilde{H}_i(\Delta, \boldsymbol{G}) = 0$ for all $i \in \mathbb{Z}$. When $\boldsymbol{G}$ is suppressed from the notation we always mean $\boldsymbol{G} = \mathbb{Z}$.

We now list some relations between homotopy properties and homology of a complex Δ, which are frequently useful. They are consequences of the theorems of Hurewicz and Whitehead (see Spanier 1966).

9.16. Δ is k-connected iff Δ is k-acyclic (over $\mathbb{Z}$) and simply connected, $k \geqslant 1$.

9.17. Δ is contractible iff Δ is $\mathbb{Z}$-acyclic and simply connected.

9.18. If Δ is simply connected, $\tilde{H}_i(\Delta) = 0$ for $i \neq d > 1$, and $\tilde{H}_d(\Delta) \cong \mathbb{Z}^k$, then Δ is homotopy equivalent to a wedge of k d-spheres.

9.19. Assume $\dim \Delta = d \geqslant 0$. Then Δ is $(d-1)$-connected iff Δ is homotopy equivalent to a wedge of d-spheres.

[Remark: The analogues of (9.17)–(9.19) may fail for non-triangulable spaces.]

9.20. If Δ_1 is k_1-acyclic and Δ_2 is k_2-acyclic then $\Delta_1 * \Delta_2$ is $(k_1 + k_2 + 2)$-acyclic. This follows from (9.12). Using (9.16) it implies that if Δ_i is k_i-connected then $\Delta_1 * \Delta_2$ is $(k_1 + k_2 + 2)$-connected. (For this, see also Milnor 1956.)

10. Combinatorial homotopy theorems

In this section we collect some tools for manipulating homotopies and the homotopy type of complexes and posets, which have proven to be useful in combinatorics. Parallel tools for homology exist in most cases. We begin with some elementary lemmas.

Suppose Δ is a simplicial complex and T a space. Let $C: \Delta \to 2^T$ be order-preserving (i.e., $C(\sigma) \subseteq C(\tau) \subseteq T$, for all $\sigma \subseteq \tau$ in Δ). A mapping $f: \|\Delta\| \to T$ is *carried* by C if $f(\|\sigma\|) \subseteq C(\sigma)$ for all $\sigma \in \Delta$. Let $k \in \mathbb{Z}_+ \cup \{\infty\}$.

Lemma 10.1 (Carrier Lemma). *Assume that $C(\sigma)$ is $\min(k, \dim(\sigma))$-connected for all $\sigma \in \Delta$. Then:*

(i) *if $f, g: \|\Delta^{\leqslant k}\| \to T$ are both carried by C, then $f \sim g$,*

(ii) *there exists a mapping $\|\Delta^{<k+1}\| \to T$ carried by C.*

In particular, if $C(\sigma)$ is always contractible then $\|\Delta\|$ can replace the skeleta in (i) and (ii) ($k = \infty$ case). Carrier lemmas of various kinds are common in topology. For proofs of this version, see Lundell and Weingram (1969) or Walker (1981b).

Lemma 10.2 (Contractible Subcomplex Lemma). *If Δ_0 is a contractible subcomplex of a simplicial complex Δ, then the projection map $\|\Delta\| \to \|\Delta\|/\|\Delta_0\|$ is a homotopy equivalence.*

This is a consequence of the homotopy extension property for simplicial pairs [for more details see Brown (1968) or Björner and Walker (1983)].

a 10.3 (Gluing Lemma). *Examples of simple gluing results for simplicial*
es Δ_1 and Δ_2 are:
and $\Delta_1 \cap \Delta_2$ are contractible, then $\Delta_1 \cup \Delta_2 \simeq \Delta_2$,
nd Δ_2 are k-connected and $\Delta_1 \cap \Delta_2$ is $(k-1)$-connected, then $\Delta_1 \cup \Delta_2$

nd $\Delta_1 \cap \Delta_2$ are k-connected, then so are also Δ_1 and Δ_2.

Such results are often special cases of the theorems in this section, especially Theorem 10.6. Otherwise they can be deduced from the Mayer–Vietoris long exact sequence (for k-acyclicity) and the Seifert–van Kampen theorem (for simply-connectedness), using (9.16) and (9.17).

A general principle for gluing homotopies appears in Brown (1968, p. 240) and Mather (1966). It gives a convenient proof for part (i) of the following lemma. For part (ii) use Lemma 10.2. A more general method for gluing homotopies (the "diagrams of spaces" technique) appears in Ziegler and Živaljević (1993).

Lemma 10.4. *Let $\Delta = \Delta_0 \cup \Delta_1 \cup \cdots \cup \Delta_n$ be a simplicial complex with subcomplexes Δ_i, and assume that $\Delta_i \cap \Delta_j \subseteq \Delta_0$ for all $1 \leqslant i < j \leqslant n$.*

(i) *If Δ_i is contractible for all $1 \leqslant i \leqslant n$, then*

$$\Delta \simeq \Delta_0 \cup \bigcup_{i=1}^{n} \operatorname{cone}(\Delta_0 \cap \Delta_i)$$

(i.e., raise a cone independently over each subcomplex $\Delta_0 \cap \Delta_i$).

(ii) *If Δ_i is contractible for all $0 \leqslant i \leqslant n$, then*

$$\Delta \simeq \operatorname{wedge}_{1 \leqslant i \leqslant n} \operatorname{susp}(\Delta_0 \cap \Delta_i).$$

Some of the following results concern *simplicial* maps $f : \Delta \to P$ from a simplicial complex Δ to a poset P. Such a map sends vertices of Δ to elements of P in such a way that each $\sigma \in \Delta$ is mapped to a chain in P. In particular, an order-preserving or order-reversing mapping of posets $Q \to P$ is of this type.

Theorem 10.5 (Fiber Theorem, Quillen 1978, Walker 1981b). *Let $f : \Delta \to P$ be a simplicial map from a simplicial complex Δ to a poset P.*

(i) *Suppose all fibers $f^{-1}(P_{\geqslant x}), x \in P$, are contractible. Then f induces homotopy equivalence between Δ and P.*

(ii) *Suppose all fibers $f^{-1}(P_{\geqslant x}), x \in P$, are k-connected. Then Δ is k-connected if and only if P is k-connected.*

Proof. Suppose that all fibers are contractible. Then the mapping $C(\sigma) = f^{-1}(P_{\geqslant \min \sigma})$, $\sigma \in \Delta(P)$, is a contractible carrier from $\Delta(P)$ to $\|\Delta\|$. By Lemma 10.1 (ii) there exists a continuous map $g : \Delta(P) \to \Delta$ carried by C, i.e., $g(\|\sigma\|) \subseteq \|f^{-1}(P_{\geqslant \min \sigma})\|$, for every chain $\sigma \in \Delta(P)$. One sees that g is a homotopy inverse to f as follows, using Lemma 10.1 (i): $C'(\sigma) = \|P_{\geqslant \min \sigma}\|, \sigma \in \Delta(P)$, is contractible and carries $f \circ g$ and id_P, and $C''(\pi) = \|f^{-1}(P_{\geqslant \min f(\pi)})\|, \pi \in \Delta$, is contractible and carries $g \circ f$ and id_Δ. Hence, $f \circ g \sim \mathrm{id}_P$ and $g \circ f \sim \mathrm{id}_\Delta$.

The second part is proved analogously by passing to $(k+1)$-skeleta and using k-connected carriers in Lemma 10.1. □

The *nerve* of a family of sets $(A_i)_{i \in I}$ is the simplicial complex $\mathcal{N} = \mathcal{N}(A_i)$ defined on the vertex set I so that a finite subset $\sigma \subseteq I$ is in $\mathcal{N}$ precisely when $\bigcap_{i \in \sigma} A_i \neq \emptyset$.

Theorem 10.6 (Nerve Theorem, Borsuk 1948, Björner et al. 1985, 1994). *Let Δ be a simplicial complex (or, a regular cell complex) and $(\Delta_i)_{i\in I}$ a family of subcomplexes such that $\Delta = \bigcup_{i\in I} \Delta_i$.*

(i) *Suppose every nonempty finite intersection $\Delta_{i_1} \cap \Delta_{i_2} \cap \cdots \cap \Delta_{i_t}$ is contractible. Then Δ and the nerve $\mathcal{N}(\Delta_i)$ are homotopy equivalent.*

(ii) *Suppose every nonempty finite intersection $\Delta_{i_1} \cap \Delta_{i_2} \cap \cdots \cap \Delta_{i_t}$ is $(k - t + 1)$-connected.Then Δ is k-connected if and only if $\mathcal{N}(\Delta_i)$ is k-connected.*

Proof. For convenience, assume that the covering of Δ by the Δ_i's is *locally finite*, meaning that each vertex of Δ belongs to only finitely many subcomplexes Δ_i. (The case of more general coverings requires a slightly different argument.)

Let $Q = P(\Delta)$ and $P = P(\mathcal{N}(\Delta_i))$ be the face posets. Define a mapping $f : Q \to P$ by $\pi \longmapsto \{i \in I \mid \pi \in \Delta_i\}$. Clearly f is order-reversing, so $f : \Delta(Q) \to P$ is simplicial. The fiber at $\sigma \in P$ is $f^{-1}(P_{\geq \sigma}) = \bigcap_{i\in\sigma} \Delta_i$. Part (i) now follows from Theorem 10.5. Also, if all nonempty finite intersections are k-connected, part (ii) follows the same way. In the stated generality, part (ii) is proved in Björner et al. (1994). □

The Nerve Theorem has several versions for coverings of a topological space by subspaces. The earliest of these seem to be due to Leray (1945) and Weil (1952). Discussions of results of this kind can be found in Wu (1962) and McCord (1967). We state here a version which seems suitable for use in combinatorics. An application to oriented matroids appears in Edelman (1984).

Theorem 10.7 (Nerve Theorem, Weil 1952, Wu 1962, McCord 1967). *Let X be a triangulable space and $(A_i)_{i\in I}$ a locally finite family of open subsets (or a finite family of closed subsets) such that $X = \bigcup_{i\in I} A_i$. If every nonempty intersection $A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_t}$ is contractible, then X and the nerve $\mathcal{N}(A_i)$ are homotopy equivalent.*

By *locally finite* is meant that each point of X lies in at most finitely many sets A_i. We warn that Theorem 10.7 is false for locally finite coverings by *closed* sets and also for too general spaces X. For a counterexample in the first case, take X to be the unit circle and $A_i = \{e^{2\pi i t} \mid 1/(i+1) \leqslant t \leqslant 1/i\}, i = 1, 2, \ldots$. In the second case one can, e.g., let X be the wedge of two topologist's combs A_1 and A_2 [as in Spanier (1966, Ex. 5, p. 56)].

The conclusions in part (ii) of Theorems 10.5 and 10.6 can be strengthened: In Theorem 10.5, if all fibers are k-connected, then f induces isomorphisms of homotopy groups $\pi_j(\Delta) \cong \pi_j(P)$, for all $j \leqslant k$. Consequently, if in Theorem 10.6 all nonempty finite intersections $\Delta_{i_1} \cap \Delta_{i_2} \cap \cdots \cap \Delta_{i_t}$ are k-connected, then $\pi_j(\Delta) \cong \pi_j(\mathcal{N}(\Delta_i))$, for all $j \leqslant k$. A similar k-connectivity version of Theorem 10.7 appears in Wu (1962).

Let P be a poset. A subset $C \subseteq P$ is called a *crosscut* if (1) C is an antichain, (2) for every finite chain σ in P there exists some element in C which is comparable to each element in σ, (3) if $A \subseteq C$ is *bounded* (here meaning that A has an upper bound or a lower bound in P) then the join $\vee A$ or the meet $\wedge A$ exists in P. For instance, the atoms of a lattice L of finite length form a crosscut in L and in $\bar{L}$.

A crosscut C in P determines the simplicial complex $\Gamma(P,C)$ consisting of the bounded subsets of C.

Theorem 10.8 (Crosscut Theorem, Rota 1964, Folkman 1966, Björner 1981). *The crosscut complex $\Gamma(P,C)$ and P are homotopy equivalent.*

Proof. For $x \in C$, let $\Delta_x = \Delta(P_{\leqslant x} \cup P_{\geqslant x})$. Then $(\Delta_x)_{x\in C}$ is a covering of $\Delta(P)$, by condition (2), and every nonempty intersection is a cone, by condition (3), and hence contractible. Since $\Gamma(P,C) = \mathcal{N}(\Delta_x)$, Theorem 10.6 implies the result. □

The neighborhood complex of a graph defined in section 4 is a special kind of nerve complex. The following result gives a special decomposition property of neighborhood complexes of bipartite graphs.

Theorem 10.9 (Bipartite Relation Theorem, Dowker 1952, Mather 1966). *Suppose $G = (V_0, V_1, E), E \subseteq V_0 \times V_1$, is a bipartite graph, and let $\Delta_i, i = 0, 1$, be the simplicial complex whose faces are all finite subsets $\sigma \subseteq V_i$ that have a common neighbor in V_{1-i}. Then Δ_0 and Δ_1 are homotopy equivalent.*

Proof. First delete any isolated vertices from G. This does not affect Δ_0 and Δ_1. Now, for every $x \in V_1$ let Δ_x consist of all finite subsets of $\{y \in V_0 \mid (y,x) \in E\}$. Then $(\Delta_x)_{x\in V_1}$ is a covering of Δ_0 with contractible nonempty intersections. The nerve of this covering is Δ_1, so Theorem 10.6 applies. □

Theorems 10.6 (i), 10.8 and 10.9 are equivalent in the sense that either one implies the other two. The following is a variation of the Fiber Theorem 10.5.

Theorem 10.10 (Ideal Relation Theorem, Quillen 1978). *Let P and Q be posets and suppose that $R \subseteq P \times Q$ is a relation such that $(x,y) \leqslant (x',y') \in R$ implies that $(x,y) \in R$. (That is, R is an order ideal in the product poset.) Suppose furthermore that $R_x = \{y \in Q \mid (x,y) \in R\}$ and $R_y = \{x \in P \mid (x,y) \in R\}$ are contractible for all $x \in P$ and $y \in Q$. Then P and Q are homotopy equivalent.*

Proof. By symmetry it suffices to show that P and R are homotopy equivalent. By Theorem 10.5 it suffices for this to show that the fiber $\pi^{-1}(P_{\geqslant x})$ is contractible for all $x \in P$, where $\pi : R \to P$ is the projection map $\pi(x,y) = x$. Let $F_x = \pi^{-1}(P_{\geqslant x}) = \{(z,y) \in R \mid z \geqslant x\}$, and let $\rho : F_x \to R_x$ be the projection $\rho(z,y) = y$. Now, $\rho^{-1}((R_x)_{\geqslant y}) = \{(z,w) \in F_x \mid w \geqslant y\} = \{(z,w) \in R \mid (z,w) \geqslant (x,y)\}$ is a cone and hence contractible, for all $y \in R_x$. So by the Fiber Theorem F_x is homotopy equivalent to R_x, which by assumption is contractible. (Remark: There is also an obvious k-connectivity version of this result.) □

Theorem 10.11 (Order Homotopy Theorem, Quillen 1978). *Let $f,g : \Delta \to P$ be simplicial maps from a simplicial complex Δ to a poset P. If $f(x) \leqslant g(x)$ for every vertex x of Δ, then f and g are homotopic.*

Proof. For each face $\sigma \in \Delta$, let $C(\sigma) = f(\sigma) \cup g(\sigma)$. The minimal element in the chain $f(\sigma)$ is below every other element in $C(\sigma)$. So the order complex of $C(\sigma)$ is a cone, and hence contractible. Since C carries both f and g, these maps are homotopic by Lemma 10.1. □

Corollary 10.12. *Let $f : P \to P$ be an order-preserving map such that $f(x) \geqslant x$ for all $x \in P$. Then f induces homotopy equivalence between P and $f(P)$.*

If also $f^2(x) = f(x)$ for all $x \in P$ (f is then called a *closure operator* on P) then $f(P)$ is a strong deformation retract of P. The hypotheses of Theorem 10.11 and Corollary 10.12 can be weakened to that $f(x)$ and $g(x)$ [resp., $f(x)$ and x] are comparable for all x.

Call a poset P *join-contractible* (*via* p), if for some element $p \in P$ the join (least upper bound) $p \vee x$ exists for all $x \in P$. Define *meet-contractible* in dual fashion.

Corollary 10.13 (Quillen 1978). *If P is join-contractible then P is contractible.*

Proof. Since $x \leqslant p \vee x \geqslant p$, for all $x \in P$, Theorem 10.11 shows that $\mathrm{id} \sim p \vee \mathrm{id} \sim p$, i.e., the identity map on P is homotopic to the constant map p. □

The following is a consequence of Corollary 10.12, and also of Theorem 10.8.

Corollary 10.14. *Let L be a lattice of finite length and A the set of its atoms. Let $J = \{\vee B \mid B \subseteq A\}$. Then $\bar{L}$ and $\bar{L} \cap J$ are homotopy equivalent.*

Proof. The mapping $f(x) = \vee(A \cap L_{\leqslant x})$ satisfies $f^2(x) = f(x) \leqslant x$ for all $x \in \bar{L}$. Now use Corollary 10.12. □

The set of complements $\mathscr{C}\mathrm{o}(z)$ of an element z in a bounded lattice L is defined in section 3. Recall that $\bar{L} = L - \{\hat{0}, \hat{1}\}$.

Theorem 10.15 (Homotopy Complementation Theorem, Björner and Walker 1983). *Let L be a bounded lattice and $z \in \bar{L}$.*

(i) *The poset $\bar{L} - \mathscr{C}\mathrm{o}(z)$ is contractible. In particular, if L is noncomplemented then $\bar{L}$ is contractible.*

(ii) *If $\mathscr{C}\mathrm{o}(z)$ is an antichain, then*

$$\bar{L} \simeq \underset{y \in \mathscr{C}\mathrm{o}(z)}{\mathrm{wedge}} \ \mathrm{susp}(\bar{L}_{<y} * \bar{L}_{>y}).$$

Proof. For each chain σ in $P = \bar{L} - \mathscr{C}\mathrm{o}(z)$, let $C(\sigma) = \{x \in P \mid x \geqslant z\} \cup \{y \in P \mid y \leqslant \max \sigma\}$. Either $z \vee \max \sigma$ exists in P, in which case $C(\sigma)$ is meet-contractible via it, or else $z \wedge \max \sigma$ exists, and $C(\sigma)$ is join-contractible via it. So, C is contractible and carries the constant map z as well as id_P. Therefore by Lemma 10.1 $z \sim \mathrm{id}_P$, which proves part (i). Part (ii) then follows by Lemma 10.4 (ii). □

Suppose that L is a bounded lattice whose proper part is not contractible. Then by part (i) every element x has a complement in L. This conclusion can be strengthened in the following way: [Lovász and Schrijver (unpublished)] Every chain $x_0 < x_1 < \cdots < x_k$ in $\bar{L}$ has a complementing chain $y_0 \geqslant y_1 \geqslant \cdots \geqslant y_k$ (i.e., $x_i \perp y_i$ for $0 \leqslant i \leqslant k$). Here one can even demand that each complement y_i is a join of atoms (assuming that atoms exist, which is the case, e.g., if L is of finite length).

A more general poset version of Theorem 10.15 is given in Björner (1994b). There the antichain assumption is dropped from part (ii) at the price of a more complicated description of the right-hand side as a quotient space of a wedge indexed by pairs $x \leqslant y$ in $\mathscr{C}\mathrm{o}(z)$.

11. Complexes with special structure

Some special properties of complexes that are frequently encountered in combinatorics, and which express a certain simplicity of structure, will be reviewed.

Collapsible and shellable complexes

11.1. Let Δ be a simplicial complex, and suppose that $\sigma \in \Delta$ is a proper face of exactly one simplex $\tau \in \Delta$. Then the complex $\Delta' = \Delta \backslash \{\sigma, \tau\}$ is obtained from Δ by an *elementary collapse* (and Δ is obtained from Δ' by an *elementary anticollapse*). Note that $\Delta' \simeq \Delta$. If Δ can be reduced to a single point by a sequence of elementary collapse steps, then Δ is *collapsible*.

The class of *nonevasive* complexes is recursively defined as follows: (i) a single vertex is nonevasive, (ii) if for some $x \in \Delta^0$ both $\mathrm{lk}_\Delta(x)$ and $\mathrm{dl}_\Delta(x)$ are nonevasive, then so is Δ.

The following logical implications are strict (i.e., converses are false):

$$\text{cone} \Longrightarrow \text{nonevasive} \Longrightarrow \text{collapsible} \Longrightarrow \text{contractible} \Longrightarrow \mathbb{Z}\text{-acyclic}.$$

Furthermore, for an arbitrary field $\boldsymbol{k}$:

$$\mathbb{Z}\text{-acyclic} \Longrightarrow \boldsymbol{k}\text{-acyclic} \Longrightarrow \mathbb{Q}\text{-acyclic} \Longrightarrow \tilde{\chi} = 0,$$

and $\mathbb{Z}$-acyclic $\Longleftrightarrow$ $\mathbb{Z}_p$-acyclic for all prime numbers p.

Nonevasive complexes were defined by Kahn et al. (1984) to model the notion of argument complexity discussed in section 2. A complex Δ is nonevasive iff for all $F \subseteq \Delta^0$ it is possible in less than card Δ^0 questions of the type "Is $x \in F$?" to decide whether $F \in \Delta$.

Collapsibility has long been studied in combinatorial topology. Noteworthy is the fact that two simply connected finite complexes Δ and Δ' are homotopy equivalent iff a sequence of elementary collapses and elementary anticollapses can transform Δ into Δ' (see Cohen 1973). In particular, the contractible complexes are precisely the complexes that collapse/anticollapse to a point.

An element x in a poset P is *irreducible* if $P_{>x}$ has a least element or $P_{<x}$ a greatest element. A finite poset is *dismantlable* if successive removal of irreducibles leads to a single-element poset. A dismantlable poset is nonevasive. A topological characterization of dismantlable posets of Stong (1966) is mentioned in (9.3). A *directed* poset (for all $x, y \in P$ there exists $z \in P$ such that $x, y \leqslant z$) is contractible.

11.2. Let Δ be a pure d-dimensional simplicial complex, and suppose that the k-face σ is contained in exactly one d-face τ. Then the complex $\Delta' = \Delta \backslash \{\gamma \mid \sigma \subseteq \gamma \subseteq \tau\}$ is obtained from Δ by a *(k,d)-collapse*. If $\sigma \neq \tau$, then $\Delta' \simeq \Delta$. If Δ can be reduced to a single d-simplex by a sequence of (k,d)-collapses, $0 \leqslant k \leqslant d$, then Δ is *shellable*.

A pure simplicial complex Δ is *vertex-decomposable* if (i) $\Delta = \emptyset$, or (ii) Δ consists of a single vertex, or (iii) for some $x \in \Delta^0$ both $\mathrm{lk}_\Delta(x)$ and $\mathrm{dl}_\Delta(x)$ are vertex-decomposable. For example, every simplex and simplex-boundary is vertex-decomposable. The class of *constructible* complexes is defined by: (i) every simplex and $\emptyset$ is constructible, (ii) if Δ_1, Δ_2 and $\Delta_1 \cap \Delta_2$ are constructible and $\dim \Delta_1 = \dim \Delta_2 = 1 + \dim(\Delta_1 \cap \Delta_2)$, then $\Delta_1 \cup \Delta_2$ is constructible.

The following logical implications between these properties of a pure d-dimensional complex are strict:

$$\text{vertex-decomposable} \Longrightarrow \text{shellable} \Longrightarrow \text{constructible}$$
$$\Longrightarrow (d-1)\text{-connected.}$$

The first implication and the definition of vertex-decomposable complexes are due to Provan and Billera (1980). The concept of shellability has an interesting history going back to the 19th century, see Grünbaum (1967). Constructible complexes were defined by M. Hochster, see Stanley (1977).

Shellability is usually regarded as a way of *putting together* (rather than collapsing – taking apart) a complex. Therefore the following alternative definition is more common: A finite pure d-dimensional complex Δ is *shellable* if its d-faces can be ordered $\sigma_1, \sigma_2, \ldots, \sigma_t$ so that $(\delta\sigma_1 \cup \cdots \cup \delta\sigma_{k-1}) \cap \delta\sigma_k$ is a pure $(d-1)$-dimensional complex for $2 \leqslant k \leqslant t$, where $\delta\sigma_j = 2^{\sigma_j} \backslash \{\emptyset, \sigma_j\}$ is the boundary complex of σ_j. Equivalently, for all $1 \leqslant i < k \leqslant t$ there exists $j < k$ such that $\sigma_i \cap \sigma_k \subseteq \sigma_j \cap \sigma_k$ and $\dim(\sigma_j \cap \sigma_k) = d - 1$. In words, the requirement is that the kth facet σ_k intersects the union of the preceding ones along a part of its boundary which is a union of maximal proper faces of σ_k. Such an ordering of the facets is called a *shelling*.

If $\sigma \in \Delta$ and Δ is a shellable (or constructible) complex, then so is $\mathrm{lk}_\Delta(\sigma)$. Shellability is also preserved by some other constructions on complexes and posets such as Theorem 11.13. Several basic properties of simplicial shellability (also for infinite complexes) are reviewed in Björner (1984b). Shellability of cell complexes is discussed in Danaraj and Klee (1974) and Björner (1984a); see also chapter 18 by Klee and Kleinschmidt. To establish shellability of (order complexes of) posets, a special method exists called *lexicographic shellability*. See Björner (1980) and Björner and Wachs (1983, 1994) for details. The notions of shellability and vertexdecomposability and most of their useful properties can easily be generalized to non-pure complexes, see Björner and Wachs (1994).

11.3. Simplicial PL spheres and PL balls are defined in (12.2), (PL = piecewise linear). The property of being PL is a combinatorial property – whether a geometric simplicial complex Δ is PL depends only on the abstract simplicial complex Δ.

For showing that specific complexes are homeomorphic to spheres or balls, the following result is frequently useful.

Theorem 11.4. *Let Δ be a constructible d-dimensional simplicial complex.*

(i) *If every $(d-1)$-face is contained in exactly two d-faces, then Δ is a PL sphere.*

(ii) *If every $(d-1)$-face is contained in one or two d-faces, and containment in only one d-face occurs, then Δ is a PL ball.*

Theorem 11.4 follows from some basic PL topology such as the facts quoted in (12.2). For shellable Δ it appears implicitly in Bing (1964) and explicitly in Danaraj and Klee (1974).

If Δ is a triangulation of the d-sphere (or any manifold) and $\sigma \in \Delta^k$, then $\mathrm{lk}_\Delta(\sigma)$ has the same homology as the $(d-1-k)$-sphere. If $\sigma \in \Delta^0$, then there is even homotopy equivalence between $\mathrm{lk}_\Delta(\sigma)$ and S^{d-1}. However, if Δ is a PL d-sphere and $\sigma \in \Delta^k$, then $\mathrm{lk}_\Delta(\sigma)$ is itself a PL $(d-1-k)$-sphere.

Cohen–Macaulay complexes

11.5. Let $\boldsymbol{k}$ be a field or the ring of integers $\mathbb{Z}$. A finite-dimensional simplicial complex Δ is *Cohen–Macaulay over $\boldsymbol{k}$* (written CM/$\boldsymbol{k}$ or CM if $\boldsymbol{k}$ is understood or irrelevant) if $\mathrm{lk}_\Delta(\sigma)$ is $(\dim \mathrm{lk}_\Delta(\sigma)-1)$-acyclic over $\boldsymbol{k}$ for all $\sigma \in \Delta \cup \{\emptyset\}$. Further, Δ is *homotopy-Cohen–Macaulay* if $\mathrm{lk}_\Delta(\sigma)$ is $(\dim \mathrm{lk}_\Delta(\sigma)-1)$-connected for all $\sigma \in \Delta \cup \{\emptyset\}$.

The following implications are strict:

$$\text{constructible} \Longrightarrow \text{homotopy-CM} \Longrightarrow \mathrm{CM}/\mathbb{Z} \Longrightarrow \mathrm{CM}/\boldsymbol{k} \Longrightarrow \mathrm{CM}/\mathbb{Q},$$

for an arbitrary field $\boldsymbol{k}$. Furthermore, CM/$\mathbb{Z} \Longleftrightarrow$ CM/$\mathbb{Z}_p$ for all prime numbers p. The first implication follows from the fact that constructibility implies $(d-1)$-connectivity and is inherited by links, the second implication follows from (9.15), and the rest via the Universal Coefficient Theorem. In particular, shellable complexes are homotopy-CM.

An important aspect of *finite* CM-complexes Δ is that they have an equivalent ring-theoretic definition. Suppose that $\Delta^0 = \{x_1, x_2, \ldots, x_n\}$, and consider the ideal I in the polynomial ring $\boldsymbol{k}[x_1, x_2, \ldots, x_n]$ generated by monomials $x_{i_1}x_{i_2}\ldots x_{i_k}$ such that $\{x_{i_1}, x_{i_2}, \ldots, x_{i_k}\} \notin \Delta, 1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n, k \geqslant 1$. Let $\boldsymbol{k}[\Delta] = \boldsymbol{k}[x_1, \ldots, x_n]/I$, called the *Stanley–Reisner ring* (or *face ring*) of Δ. Then Δ is CM/$\boldsymbol{k}$ iff the ring $\boldsymbol{k}[\Delta]$ is Cohen–Macaulay in the sense of commutative algebra (Reisner 1976). An exposition of the ring-theoretic aspects of simplicial complexes, and their combinatorial use, can be found in Stanley (1983a). There other ring-theoretically motivated classes of complexes, such as *Gorenstein complexes* and *Buchsbaum complexes*, are also discussed. Other approaches to the ring-theoretic aspects of complexes and to Reisner's theorem can be found in Baclawski and Garsia (1981) and Yuzvinsky (1987). See also section 5 of chapter 41 on Combinatorics in Pure Mathematics.

Cohen–Macaulay complexes and posets were introduced around 1974–75 in the work of Baclawski (1976, 1980), Hochster (1977), Reisner (1976) and Stanley (1975, 1977). The notion of homotopy-CM first appeared in Quillen (1978). Björner, Garsia and Stanley (1982) give an elementary introduction to CM posets. A notable combinatorial application of Cohen–Macaulayness is Stanley's proof of tight upper bounds for the number of faces that can occur in each dimension for triangulations with n vertices of the d-sphere (Stanley 1975, 1983a; see also chapter 18 by Klee and Kleinschmidt.) An application to lower bounds is given in Stanley (1987a).

11.6. Define a pure d-dimensional complex Δ to be *strongly connected* (or *dually connected*) if each pair of facets $\sigma, \tau \in \Delta^d$ can be connected by a sequence of facets $\sigma = \sigma_0, \sigma_1, \ldots, \sigma_n = \tau$, so that $\dim(\sigma_{i-1} \cap \sigma_i) = d - 1$ for $1 \leqslant i \leqslant n$.

Proposition 11.7. *Every CM complex is pure and strongly connected.*

This follows from the following lemma, which is proved by induction on $\dim \Delta$: *Let Δ be a finite-dimensional simplicial complex, and assume that* $\mathrm{lk}_\Delta(\sigma)$ *is connected for all $\sigma \in \Delta \cup \{\emptyset\}$ such that* $\dim(\mathrm{lk}_\Delta(\sigma)) \geqslant 1$. *Then Δ is pure and strongly connected.*

The property of being CM is *topologically invariant*: whether Δ is CM/$\boldsymbol{k}$ or not depends only on the topology of $\|\Delta\|$. This is implied by the following reformulation of CM-ness, due to Munkres (1984b).

Theorem 11.8. *A finite-dimensional complex Δ is* CM/$\boldsymbol{k}$ *iff its space $T = \|\Delta\|$ satisfies: $\tilde{H}_i(T, \boldsymbol{k}) = H_i(T, T\backslash p, \boldsymbol{k}) = 0$ for all $p \in T$ and $i <$* $\dim \Delta$.

In this formulation $\tilde{H}_i$ denotes reduced singular homology and H_i relative singular homology with coefficients in $\boldsymbol{k}$. A consequence of Theorem 11.8 is that if M is a triangulable manifold (with or without boundary) and $\tilde{H}_i(M) = 0$ for $i < \dim M$, then every triangulation of M is CM. For instance: (1) every triangulation of the d-sphere, d-ball or $\mathbb{R}^d$ is CM/$\mathbb{Z}$, but not necessarily homotopy-CM (beware: homotopy-CM is *not* topologically invariant), (2) a triangulation of real projective d-space is CM/$\boldsymbol{k}$ iff char $\boldsymbol{k} \neq 2$.

11.9. The definition of Cohen–Macaulay posets (posets P such that $\Delta(P)$ is CM) deserves a small additional comment. Let P be a poset of finite rank and σ: $x_0 < x_1 < \cdots < x_k$ a chain in P. Then $\mathrm{lk}_{\Delta(P)}(\sigma) = P_{<x_0} * (x_0, x_1) * \cdots * (x_{k-1}, x_k) * P_{>x_k}$. It therefore follows from (9.20) that P is CM [resp. homotopy-CM] iff every open interval (x, y) in $\hat{P}$ is $(\mathrm{rank}(x,y) - 1)$-acyclic [resp. $(\mathrm{rank}(x,y) - 1)$-connected].

Some uses of Cohen–Macaulay posets in commutative algebra are discussed in section 5 of chapter 41 on Combinatorics in Pure Mathematics.

11.10. An abundance of shellable and CM simplicial complexes appear in combinatorics. Only a few important examples can be mentioned here.

(i) The boundary complex of a simplicial convex polytope is shellable (Bruggesser and Mani 1971, Danaraj and Klee 1974; see also chapter 18 by Klee and Kleinschmidt). Every simplicial PL sphere is the boundary of a shellable ball (Pachner 1986). There exist non-shellable triangulations of the 3-ball (M.E. Rudin) and of the 3-sphere (see below). Shellability of spheres and balls is surveyed in Danaraj and Klee (1978).

(ii) The following implications are valid for any simplicial sphere: constructible $\Longrightarrow PL \Longrightarrow$ homotopy-CM. The 5-sphere admits triangulations that are non-homotopy-CM (R.D. Edwards, see Daverman 1986), and also PL triangulations that are non-constructihle (Mandel 1982). Every triangulation of the 3-sphere is PL, but all are not shellable (Lickorish 1991, see also Vince 1985). Face lattices of regular complex polytopes are CM (Orlik 1990).

(iii) The complex of independent sets in a matroid is constructible (Stanley 1977) and vertex-decomposable (Provan and Billera 1980). More generally, the complex generated by the basis-complements of a greedoid is vertex-decomposable (Björner, Korte and Lovász 1985). Complexes arising from matroids are discussed in Björner (1992).

(iv) Every semimodular (in particular, every geometric or modular) lattice of finite rank is CM (Folkman 1966) and shellable (Björner 1980). For any element $x \neq \hat{0}$ in a geometric lattice L, the poset $L \backslash [x, \hat{1}]$ is shellable (Wachs and Walker 1986).

(v) Tits buildings are CM (Solomon–Tits, see Brown 1989 or Ronan 1989) and shellable (Björner 1984b). The topology of more general group-related geometries has been studied by Ronan (1981), Smith (1988), Tits (1981) and others with a view to uses in group theory. See Buekenhout (1995) and Ronan (1989) for general accounts.

(vi) The poset of elementary Abelian p-subgroups of a finite group was shown by Quillen (1978) to be homotopy-CM in some cases. See also Stong (1984). The full subgroup lattice of a finite group G is shellable (or CM) iff G is supersolvable (Björner 1980). Various posets of subgroups have been studied from a topological point of view. See Thévenaz (1987), Webb (1987) and Welker (1994) for a guide to this literature.

Induced subcomplexes

Connectivity, Cohen–Macaulayness, etc., are under certain circumstances inherited by suitable subcomplexes. For a simplicial complex Δ and $A \subseteq \Delta^0$, let $\Delta_A = \{\sigma \in \Delta \mid \sigma \subseteq A\}$ (the *induced subcomplex* on A).

Lemma 11.11. *Let Δ be a finite-dimensional complex, and $A \subseteq V = \Delta^0$. Assume that $\mathrm{lk}_\Delta(\sigma)$ is k-connected for all $\sigma \in \Delta_{V \backslash A}$. Then Δ_A is k-connected iff Δ is k-connected.*

Lemma 11.12. *Let P be a poset of finite rank and A a subset. Assume that $P_{>x}$ is k-connected for all $x \in P \backslash A$. Then A is k-connected iff P is k-connected.*

Proof. These lemmas are equivalent. We start with Lemma 11.12. Let $f: A \to P$ be the embedding map. For $x \in P$,

$$f^{-1}(P_{\geqslant x}) = \begin{cases} A_{\geqslant x}, & \text{if } x \in A, \\ P_{>x} \cap A, & \text{if } x \notin A. \end{cases}$$

Now, $A_{\geqslant x}$ is contractible (being a cone), and $P_{>x} \cap A$ is k-connected by induction on rank(P). The result therefore follows by Theorem 10.5 (ii).

To prove Lemma 11.11, let $P = P(\Delta)$ and $Q = \{\tau \in \Delta \mid \tau \cap A \neq \emptyset\} \subseteq P$. Since $P_{>\sigma} \cong P(\mathrm{lk}_\Delta(\sigma))$ is k-connected for all $\sigma \in P \backslash Q$, Lemma 11.12 applies. On the other hand, by Corollary 10.12 the map $f(\tau) = \tau \cap A$ on Q induces homotopy equivalence between Q and $f(Q) = P(\Delta_A)$. □

The homology versions of Lemmas 11.11 and 11.12, obtained by using k-acyclicity throughout, can be proven by a parallel method. Also, if the hypothesis "k-connected" were replaced by "contractible" in these lemmas, then the conclusion would be that Δ_A and Δ (resp. A and P) are homotopy equivalent.

Theorem 11.13. *Let Δ be a pure d-dimensional simplicial complex, $A \subseteq \Delta^0$ and $1 \leqslant m \leqslant d$. Suppose that* card$(A \cap \sigma) = m$ *for every facet $\sigma \in \Delta^d$. If Δ is* CM/$\boldsymbol{k}$, *homotopy-*CM *or shellable, then the same property is inherited by Δ_A.*

For CM-ness this result was proven in varying degrees of generality by Baclawski (1980), Munkres (1984b), Stanley (1979) and Walker (1981a). It follows easily from Lemma 11.11. For shellability, proofs appear in Björner (1980, 1984b).

Suppose that Δ is a pure d-dimensional simplicial complex and that there exists a mapping $t: \Delta^0 \to \{0, 1, \ldots, d\}$ which restricts to a bijection on each facet $\sigma \in \Delta^d$. Then Δ is called *completely balanced* (or *numbered*, or *colored*) with *type-map t*. For instance, the order complex of a pure poset is completely balanced with type-map $t = rank$ [cf. (9.2)], and also building-like incidence geometries (Buekenhout 1995) give rise to completely balanced complexes. CM complexes of this kind were studied by Stanley (1979) and others.

For each $J \subseteq \{0, 1, \ldots, d\}$, the *type-selected subcomplex* $\Delta_{(J)} = \Delta_{t^{-1}(J)}$ is the induced subcomplex on $t^{-1}(J) \subseteq \Delta^0$. Theorem 11.13 shows that if Δ is CM then $\Delta_{(J)}$ is also CM and hence (card $J - 2$)-acyclic. A certain converse is also true in the sense of the following result, which gives an alternative characterization of the CM property for completely balanced complexes. It is due to Baclawski and Garsia (1981) in the finite CM case, and to J. Walker (letter to the author, 1981) in general including the homotopy case.

Theorem 11.14. *Let Δ be a pure d-dimensional completely balanced complex. Then Δ is* CM/$\boldsymbol{k}$ *[resp., homotopy-CM] if and only if $\Delta_{(J)}$ is* (card $J - 2$)*-acyclic over* $\boldsymbol{k}$ *[resp.,* (card $J - 2$)*-connected] for all $J \subseteq \{0, 1, \ldots, d\}$.*

12. Cell complexes

Most classes of cell complexes differ from the simplicial case in that a purely combinatorial description of these objects *as such* cannot be given. However, the two classes defined here, polyhedral complexes and regular CW complexes, are sufficiently close to the simplicial case to allow a similar combinatorial approach in many cases. For simplicity only *finite* complexes will be considered.

Good general references for polyhedral complexes are Grünbaum (1967) and Hudson (1969), and for cell complexes Cooke and Finney (1967) and Lundell and Weingram (1969). Cell complexes are also discussed in many books on algebraic topology such as Munkres (1984a) and Spanier (1966).

Polyhedral complexes and PL topology

12.1. A *convex polytope* π is a bounded subset of $\mathbb{R}^d$ which is the solution set of a finite number of linear equalities and inequalities. Any nonempty subset obtained by changing some of the inequalities to equalities is a *face* of π. Equivalently, $\pi \subseteq \mathbb{R}^d$ is a convex polytope iff π is the convex hull of a finite set of points in $\mathbb{R}^d$. See chapter 18 by Klee and Kleinschmidt for more information about convex polytopes.

A *polyhedral complex* (or *convex cell complex*) Γ is a finite collection of convex polytopes in $\mathbb{R}^d$ such that (i) if $\pi \in \Gamma$ and σ is a face of π then $\sigma \in \Gamma$, and (ii) if $\pi, \tau \in \Gamma$ and $\pi \cap \tau \neq \emptyset$ then $\pi \cap \tau$ is a face of both π and τ. The members of Γ are called *cells*. The *underlying space* of Γ is $\|\Gamma\| = \bigcup \Gamma$, with the topology induced as a subset of $\mathbb{R}^d$. If every cell in Γ is a *simplex* (the convex hull of an affinely independent set of points) then Γ is called a (*geometric*) *simplicial complex*. The *dimension* of a cell equals the linear dimension of its affine span, and $\dim \Gamma = \max_{\pi \in \Gamma} \dim \pi$. Further terminology, such as *vertices, edges, facets, pure, k-skeleton, face poset, face lattice*, etc., is defined just as in the simplicial case, see (9.1) and (9.3).

12.2. A polyhedral complex Γ_1 is a *subdivision* of another such complex Γ_2 if $\|\Gamma_1\| = \|\Gamma_2\|$ and every cell of Γ_1 is a subset of some cell of Γ_2. The abstract simplicial complex $\Delta(P(\Gamma))$, i.e., the order complex of Γ's face poset, has geometric realizations (by choosing as new vertices an interior point in each cell) that subdivide Γ. Every polyhedral complex can be simplicially subdivided without introducing new vertices.

Let Σ^d denote the complex consisting of a geometric d-simplex and all its faces, and let $\delta\Sigma^d$ denote its boundary. These complexes provide the simplest triangulations of the d-ball and the $(d-1)$-sphere, respectively. A polyhedral complex Γ is called a PL *d-ball* (or PL $(d-1)$*-sphere*) if it admits a simplicial subdivision whose face poset is isomorphic to the face poset of some subdivision of Σ^d (resp. $\delta\Sigma^d$). This is equivalent to saying that there exists a homeomorphism $\|\Gamma\| \to \|\Sigma^d\|$ (resp.

$\|\Gamma\| \to \|\delta\Sigma^d\|$) which is induced by a simplicial map defined on some subdivision (a *piecewise linear*, or PL, *map*). The boundary complex of a convex d-polytope is a PL $(d-1)$-sphere.

The PL property is mainly of technical interest. Several properties of balls and spheres that are desirable, and would in many cases seem intuitively "obvious", hold only in the PL case. Some examples are: (1) (Newman's Theorem) *the closure of the complement of a PL d-ball lying in a PL d-sphere is itself a PL d-ball*; (2) *the union of two PL d-balls, whose intersection is a PL* $(d-1)$*-ball lying in the boundary of each, is a PL d-ball*; (3) *the link of any face in a PL sphere is itself a PL sphere* (cf. remark following Theorem 11.4). All these statements would be false with "PL" removed.

See Hudson (1969) for proofs and further information about PL topology. Mandel (1982) develops basic PL topology from a combinatorial perspective.

Regular cell complexes

12.3. By "cell complex" we will here understand what in topology is usually called a "finite CW complex".

Let X be a Hausdorff space. A subset σ is called an *open d-cell* if there exists a mapping $f : \boldsymbol{B}^d \to X$ whose restriction to the interior of the d-ball is a homeomorphism $f : \operatorname{Int}(\boldsymbol{B}^d) \to \sigma$. The *dimension* $\dim \sigma = d$ is well-defined by this. The closure $\bar{\sigma}$ is the corresponding *closed cell*. It is true that $f(\boldsymbol{B}^d) = \bar{\sigma}$, but $\bar{\sigma}$ is not necessarily homeomorphic to $\boldsymbol{B}^d$. We write $\dot{\sigma} = \bar{\sigma} \backslash \sigma$.

A *cell complex* $\mathscr{C}$ is a finite collection of pairwise disjoint sets together with a Hausdorff topology on their union $\|\mathscr{C}\| = \bigcup \mathscr{C}$ such that:

(i) each $\sigma \in \mathscr{C}$ is an open cell in $\|\mathscr{C}\|$, and

(ii) $\dot{\sigma} \subseteq \mathscr{C}^{<\dim \sigma}$ (the union of all cells in $\mathscr{C}$ of dimension less than $\dim \sigma$), for all $\sigma \in \mathscr{C}$.

Then $\mathscr{C}$ is also called a *cell decomposition* of the space $\|\mathscr{C}\|$. Furthermore, $\mathscr{C}$ is *regular* if each mapping $f : \boldsymbol{B}^d \to \|\mathscr{C}\|$ defining the cells can be chosen to be a homeomorphism on *all* of $\boldsymbol{B}^d$. Then, of course, every closed cell $\bar{\sigma}$ is homeomorphic to a ball. (However, it is not enough for the definition of a regular complex to only require that every closed cell is homeomorphic to a ball. The smallest example showing this has three vertices, three edges and one 2-cell.)

The cell decomposition of the d-sphere into one 0-cell and one d-cell (a point and its complement in $\boldsymbol{S}^d$) is not regular. Every polyhedral complex is a regular cell complex (the relative interiors of the convex polytopes are the open cells). Regular cell complexes are more general than polyhedral complexes in several ways. For instance, it is allowed that the intersection of two closed cells can have nontrivial topological structure.

12.4. From now on only regular cell complexes will be considered. Define the *face poset* $P(\mathscr{C})$ as the set of all closed cells ordered by containment.The following two

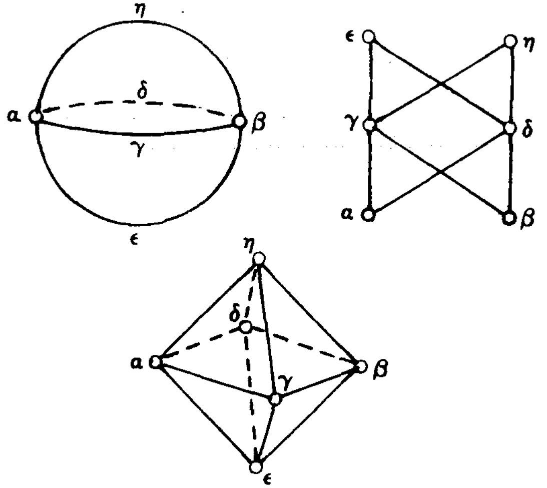


Figure 2.

particular properties make a regular complex $\mathscr{C}$ favorable from a combinatorial point of view (see Cooke and Finney 1967 or Lundell and Weingram 1969 for proofs):

(i) *The boundary $\dot{\sigma}$ of each cell $\sigma \in \mathscr{C}$ is a union of cells (a subcomplex).* Hence, the situation resembles that of polyhedral complexes: each closed d-cell $\bar{\sigma}$ is homeomorphic to $\boldsymbol{B}^d$, and its boundary $\dot{\sigma}$ (homeomorphic to S^{d-1}) has a regular cell decomposition provided by the cells that intersect $\dot{\sigma}$.

(ii) $\|\mathscr{C}\| \cong \|\Delta(P(\mathscr{C}))\|$, i.e., *the order complex of $P(\mathscr{C})$ is homeomorphic to $\|\mathscr{C}\|$.* Geometrically this means that regular cell complexes admit "barycentric subdivisions". From a combinatorial point of view it means that regular cell complexes can be interpreted as a class of posets without any loss of topological information.

Because of (i), regular cell complexes can be characterized in the following way: A family of balls (homeomorphs of $\boldsymbol{B}^d, d \geqslant 0$) in a Hausdorff space X is the set of closed cells of a regular cell complex iff the interiors of the balls partition X and the boundary of each ball is a union of other balls. This is what Mandel (1982) calls a "ball complex".

An important consequence of (ii) is that a d-dimensional regular cell complex $\mathscr{C}$ can always be "realized" in $\mathbb{R}^{2d+1}$ by a simplicial complex, so that every closed cell in $\mathscr{C}$ is a triangulated ball (a cone over a simplicial sphere).

For a detailed discussion of regular cell complexes from a combinatorial point of view, see section 4.7 of Björner et al. (1993). Figure 2 shows a regular cell decomposition $\mathscr{C}$ of the 2-sphere, its face poset $P(\mathscr{C})$, and its simplicial representation $\Delta(P(\mathscr{C}))$, where each original 2-cell is triangulated into four triangles.

12.5. Given a finite poset P, does there exist a regular cell complex (or even a polyhedral complex) $\mathscr{C}$ such that $P \cong P(\mathscr{C})$; and if so, what is its topology and how can $\mathscr{C}$ be constructed from P? This question is discussed in Björner (1984a) and Mandel (1982) from different perspectives. One answer is that P is isomorphic to the face poset of some regular cell complex iff $\Delta(P_{<x})$ is homeomorphic to a sphere for all $x \in P$. However, since it is known that simplicial spheres cannot be recognized algorithmically this is not a fully satisfactory answer. The question of how to recognize the face posets of polyhedral complexes is one version of the Steinitz problem (see chapter 18 by Klee and Kleinschmidt).

For the cellular interpretation of posets the following result, derivable from Theorem 11.4, has proven useful in practice. See Björner (1984a) for further details. Let us call a poset P *thin* if every closed interval of rank 2 has four elements (two "in the middle"). Also, $P \cup \{\hat{0}\}$ will denote P with a new minimum element $\hat{0}$ adjoined, and $\hat{P} = P \cup \{\hat{0}, \hat{1}\}$ as usual.

Theorem 12.6. *Let P be a pure finite poset of rank d. Assume that $\Delta(P)$ is constructible.*

(i) *If $P \cup \{\hat{0}\}$ is thin, then $P \cong P(\mathscr{C})$ for some regular cell complex $\mathscr{C}$ homotopy equivalent to a wedge of d-spheres.*

(ii) *If $\hat{P}$ is thin, then $P \cong P(\mathscr{C})$ for some regular cell decomposition of the d-sphere.*

13. Fixed-point and antipodality theorems

The topological fixed-point and antipodality theorems of greatest use for combinatorics will be reviewed. We start by stating four equivalent versions of the oldest of them: Brouwer's fixed-point theorem (from 1912). Proofs and references to original sources for all otherwise unreferenced material in this section can be found in many topology books, e.g., in Dugundji and Granas (1982). Recall that mappings between topological spaces are always assumed to be continuous.

Theorem 13.1 (Brouwer's Theorem). (i) *Every mapping $f: \boldsymbol{B}^d \to \boldsymbol{B}^d$ has a fixed point $x = f(x)$.*

(ii) *$\boldsymbol{S}^{d-1}$ is not a retract of $\boldsymbol{B}^d$ (i.e., no mapping $\boldsymbol{B}^d \to \boldsymbol{S}^{d-1}$ leaves each point of $\boldsymbol{S}^{d-1}$ fixed).*

(iii) *$\boldsymbol{S}^{d-1}$ is not $(d-1)$-connected.*

(iv) *$\boldsymbol{S}^{d-1}$ is not contractible.*

Brouwer's Theorem is implied by the following combinatorial lemma of Sperner (1928), see also Cohen (1967): *If the vertices of a triangulation of $\boldsymbol{S}^{d-1}$ are colored with d colors, then there cannot be exactly one $(d-1)$-face whose vertices use all d colors.* Sperner's Lemma was generalized by Lovász (1980): *If the vertices of a $(d-1)$-dimensional manifold are labeled by elements from some rank-d loopless matroid, then there cannot be exactly one $(d-1)$-face whose vertices form a basis* *of the matroid.* A further generalization and an application to hypergraphs appear in Lindström (1981). Sperner's Lemma is of practical use for the design of fixed-point-finding algorithms in connection with applications of Brouwer's Theorem, see Todd (1976).

It is well known that Brouwer's Theorem for $d = 2$ implies that there is no draw in the 2-person game HEX. Actually the implication goes the other way as well. Gale (1979) defines a d-person d-dimensional HEX game, and proves that for each $d \geqslant 2$ the Brouwer Theorem 13.1 is equivalent to the impossibility of a draw in d-dimensional HEX.

We turn next to the (Hopf–)Lefschetz fixed-point theorem (from 1927–28), which gives a vast generalization of Theorem 13.1. Lefschetz' Theorem and the closely related trace formula of Hopf will be stated in simplicial versions.

Let Δ be a nonempty simplicial complex and $f : \|\Delta\| \to \|\Delta\|$ a continuous map. The *Lefschetz number* $\Lambda(f)$ is defined by $\Lambda(f) = \sum_{i>0}(-1)^i$ trace (f_i^*), where $f_i^* : \tilde{H}_i(\Delta, \mathbb{Q}) \to \tilde{H}_i(\Delta, \mathbb{Q})$ is the induced mapping on i-dimensional reduced homology. (We use $\mathbb{Q}$-coefficients throughout here for simplicity; other fields may of course be used instead.) Note that $f \sim g$ implies $\Lambda(f) = \Lambda(g)$ (since homotopic maps induce identical maps on homology), in particular if f is *null-homotopic* (meaning homotopic to a constant map) then $\Lambda(f) = 0$. Also, if Δ is $\mathbb{Q}$-acylic then $\Lambda(f) = 0$ for all self-maps f.

Now, suppose that $f : \Delta \to \Delta$ is simplicial, and say that a face $\tau \in \Delta$ is *fixed* if $f(\tau) = \tau$ as a set. Let $\varphi_i^+(f)$ [resp. $\varphi_i^-(f)$]be the number of fixed i-faces whose orientation is preserved [resp. reversed]. Here we consider the orientation of $\tau = \{x_0, x_1, \ldots, x_i\}$ to be preserved if the permutations $x_0, x_1, \ldots, x_i$ and $f(x_0), f(x_1), \ldots, f(x_i)$ have the same parity. The following is a special case of the *Hopf trace formula*:

$$\Lambda(f) + 1 = \sum_{i \geqslant 0} (-1)^i \, [\varphi_i^+(f) - \varphi_i^-(f)]. \tag{13.2}$$

Notice that for $f = \mathrm{id}$ formula (13.2) specializes to the Euler–Poincaré formula (9.13).

One sees from (13.2) that if f has no fixed face, then $\Lambda(f) = -1$. Using simplicial approximation and compactness the following is deduced.

Theorem 13.3 (Lefschetz's Theorem). *If* $f : \|\Delta\| \to \|\Delta\|$ *is a mapping such that* $\Lambda(f) \neq -1$, *then f has a fixed point.*

The following two consequences of Theorem 13.3 generalize Brouwer's Theorem in different directions.

Corollary 13.4. *Let T be a compact triangulable space.*

(a) *Every null-homotopic self-map of T has a fixed point.*

(b) *If T is $\mathbb{Q}$-acyclic, then every self-map of T has a fixed point.*

The following consequence of the Hopf trace formula is useful in some combinatorial situations. Let once more $f : \Delta \to \Delta$ be a simplicial mapping of a simplicial complex Δ. Assume that a face $\tau \in \Delta$ is fixed if and only if τ is *point-wise* fixed [i.e., $f(\tau) = \tau$ implies $f(x) = x$ for all $x \in \tau$]. One may then define the *fixed subcomplex* $\Delta^f = \{\tau \in \Delta \mid f(\tau) = \tau\}$, which coincides with the induced subcomplex on the set of fixed vertices, and (13.2) specializes to

$$\Lambda(f) = \tilde{\chi}(\Delta^f). \tag{13.5}$$

One situation where this is used (see, e.g., Curtis, Lehrer and Tits 1980) is in connection with groups acting on finite complexes, where (13.5) says that the "Lefschetz character" has a topological interpretation as the reduced Euler characteristic of the fixed subcomplex. Another such situation (see Baclawski and Björner 1979 and section 3 of this chapter) is when $f: P \to P$ is an order-preserving poset map, in which case (13.5) can be rewritten $\Lambda(f) = \mu(P^f)$, the right-hand side denoting the value of the Möbius function computed over the subposet of fixed points augmented with a new $\hat{0}$ and $\hat{1}$ [cf. (9.14)].

The following definitions will now be needed. Let p be a prime. By a *$\mathbb{Z}_p$-space* we understand a pair (T, ν) where T is a topological space and $\nu : T \to T$ is a fixed-point free continuous mapping of order p (i.e., $\nu^p = \mathrm{id}$). A mapping $f: T_1 \to T_2$ of $\mathbb{Z}_p$-spaces $(T_i, \nu_i), i = 1,2$, is *equivariant* if $\nu_2 \circ f = f \circ \nu_1$. A $\mathbb{Z}_2$-space is often called an *antipodality space*. The standard example is (S^d, α), the d-sphere with its antipodal map $\alpha(x) = -x$.

We state five equivalent versions of the antipodality theorem of Borsuk (1933).

Theorem 13.6 (Borsuk's Theorem).

(i) *If S^d is covered by $d + 1$ subsets, all closed or all open, then one of these must contain a pair of antipodal points.* (Borsuk–Liusternik–Schnirelman)

(ii) *For every continuous mapping $f: S^d \to \mathbb{R}^d$ there exists a point x such that $f(x) = f(-x)$.* (Borsuk–Ulam)

(iii) *For every odd $[f(-y) = -f(y)$ for all $y]$ continuous mapping $f: S^d \to \mathbb{R}^d$ there exists x for which $f(x) = 0$.* (Borsuk–Ulam)

(iv) *There exists no equivariant map $S^n \to S^d$, if $n > d$.*

(v) *For any d-connected antipodality space T, there exists no equivariant map $T \to S^d$.*

Borsuk's Theorem is implied by a certain combinatorial lemma of A.W. Tucker, much like Brouwer's Theorem is implied by Sperner's Lemma. See Freund and Todd (1981) for a statement and proof of Tucker's Lemma and further references. In Theorem 13.6 (v) it suffices to assume that T is d-acyclic over $\mathbb{Z}_2$, see Walker (1983b).

Steinlein (1985) gives an extensive survey of generalizations, applications and references related to Borsuk's Theorem. Applications to combinatorics are surveyed by Alon (1988), Bárány (1993) and Bogatyi (1986); see also sections 4 and 5 of this chapter.

The following extension of the Borsuk–Ulam Theorem appears in Yang (1955): *For every mapping $S^{dn} \to \mathbb{R}^d$ there exist n mutually orthogonal diameters whose $2n$*

endpoints are mapped to the same point. The same paper also gives references to the following related theorem of Kakutani–Yamabe–Yujobô: *For every mapping $S^n \to \mathbb{R}$ there exist $(n+1)$ mutually orthogonal radii whose $(n+1)$ endpoints are mapped to the same point.* An interesting consequence of the last result is that every compact convex body $K \subset \mathbb{R}^{n+1}$ is contained in an $(n+1)$-cube C such that every maximal face of C touches K [for each $x \in S^n$ let $f(x)$ be the minimal distance between two parallel hyperplanes orthogonal to the vector x and containing K between them].

Suppose E_1 and E_2 are two bounded and measurable subsets of $\mathbb{R}^2$. Identify $\mathbb{R}^2$ with the affine plane $A = \{(\xi, \eta, 1)\}$ in $\mathbb{R}^3$, and for each $x \in S^2$ let $f_i(x)$ be the measure of that part of E_i which lies on the same side as x of the plane H_x through the origin orthogonal to x, for $i = 1, 2$. The Borsuk–Ulam Theorem implies that $f_1(x) = f_1(-x)$ and $f_2(x) = f_2(-x)$ for some $x \in S^2$, which means that the line $A \cap H_x$ bisects both E_1 and E_2. This "ham sandwich" argument generalizes to arbitrary dimensions and leads to the following consequence of the Borsuk–Ulam Theorem.

Corollary 13.7 ("Ham Sandwich Theorem"). *Given d bounded and Lebesgue measurable sets in $\mathbb{R}^d$ there exists some affine hyperplane that simultaneously bisects them all.*

Also Corollary 13.7 has several generalizations and related results. The case when $k \leqslant d$ bounded and measurable sets are given is covered by the following result of Živaljević and Vrećica (1990): *Let $\mu_1, \mu_2, \ldots, \mu_k$ be a collection of σ-additive probability measures defined on the σ-algebra of all Borel sets in $\mathbb{R}^d$, $1 \leqslant k \leqslant d$. Then there exists a $(k-1)$-dimensional affine subspace $A \subseteq \mathbb{R}^d$ such that for every closed halfspace $H \subseteq \mathbb{R}^d$ and every $i = 1, 2, \ldots, k$, $A \subseteq H$ implies $\mu_i(H) \geqslant 1/(d-k+2)$.* For $k = d$ this specializes to a measure-theoretic version of the Ham Sandwich Theorem (see also Hill 1988), and for $k = 1$ it gives a theorem of Rado (1946) which says that for any measurable $E \subseteq \mathbb{R}^d$ there exists a point $x \in \mathbb{R}^d$ such that every halfspace containing x contains at least a $1/(d+1)$-fraction of E.

We end by stating a useful generalization of the Borsuk–Ulam Theorem to $\mathbb{Z}_p$-spaces for $p > 2$. First a few definitions, see Bárány et al. (1981) for complete details. Let p be a prime and $n \geqslant 1$. Take p disjoint copies of the $n(p-1)$-dimensional ball and identify their boundaries. Call this space $X_{n,p}$. There exists a mapping $\nu : S^{n(p-1)-1} \to S^{n(p-1)-1}$ of the identified boundary which makes it into a $\mathbb{Z}_p$-space. Extend this mapping to $X_{n,p}$ as follows. If (y, r, q) denotes the point of $X_{n,p}$ from the qth ball with radius r and $S^{n(p-1)-1}$-coordinate y, then put $\nu(y, r, q) = (\nu y, r, q+1)$, where $q+1$ is reduced modulo p. This mapping ν makes $X_{n,p}$ into a $\mathbb{Z}_p$- space. [Note that $(X_{n,2}, \nu) \cong (S^n, \alpha)$.]

Theorem 13.8 (Bárány, Shlosman and Szűcs 1981). *For every continuous mapping $f \colon X_{n,p} \to \mathbb{R}^n$ there exists a point x such that $f(x) = f(\nu x) = \cdots = f(\nu^{p-1} x)$.*

Some applications of Theorem 13.8 are mentioned in sections 4 and 5.

References

Akiyama, J., and N. Alon

[1989] Disjoint simplices and geometric hypergraphs, in: *Combinatorial Mathematics – Proc. 3rd Int. Conf., 1985*, eds. G. Bloom et al., *Ann. New York Acad. Sci.* **555**, 1–3.

Aleksandrov, P.S.

[1937] Diskrete Räume, *Mat. Sbornik (N.S.)* **2**, 501–518.

Alon, N.

[1987] Splitting necklaces, *Adv. in Math.* **63**, 247–253.

[1988] Some recent combinatorial applications of Borsuk-type theorems, in: *Algebraic, Extremal and Metric Combinatorics*, eds. M. Deza, P. Frankl and D.G. Rosenberg (Cambridge University Press, Cambridge) pp. 1–12.

Alon, N., and D.B. West

[1986] The Borsuk–Ulam theorem and bisection of necklaces, *Proc. Amer. Math. Soc.* **98**, 623–628.

Alon, N., P. Frankl and L. Lovász

[1986] The chromatic number of Kneser hypergraphs, *Trans. Amer. Math. Soc.* **298**, 359–370.

Bachem, A., and W. Kern

[1992] *Linear Programming Duality – An Introduction to Oriented Matroids* (Springer, Berlin).

Bacławski, K.

[1975] Whitney numbers of geometric lattices, *Adv. in Math.* **16**, 125–138.

[1976] *Homology and combinatorics of ordered sets*, Ph.D. Thesis (Harvard University).

[1980] Cohen–Macaulay ordered sets, *J. Algebra* **63**, 226–258.

Bacławski, K., and A. Björner

[1979] Fixed points in partially ordered sets, *Adv. in Math.* **31**, 263–287.

[1981] Fixed points and complements in finite lattices, *J. Combin. Theory A* **30**, 335–338.

Bacławski, K., and A.M. Garsia

[1981] Combinatorial decompositions of a class of rings, *Adv. in Math.* **39**, 155–184.

Bajmóczy, E.G., and I. Bárány

[1979] On a common generalization of Borsuk's and Radon's theorem, *Acta Math. Acad. Sci. Hungar.* **34**, 347–350.

Bárány, I.

[1978] A short proof of Kneser's conjecture, *J. Combin. Theory A* **25**, 325–326.

[1993] Geometric and combinatorial applications of Borsuk's theorem, A survey, in: *New Trends in Discrete and Conputational Geometry*, ed. J. Pach (Springer, Berlin) pp. 235–249.

Bárány, I., and L. Lovász

[1982] Borsuk's theorem and the number of facets of centrally symmetric polytopes, *Acta Math. Acad. Sci. Hungar.* **40**, 323–329.

Bárány, I., S.B. Shlosman and A. Szűcs

[1981] On a topological generalization of a theorem of Tverberg, *J. London Math. Soc. (2)* **23**, 158–164.

Bayer, M.M., and C.W. Lee

[1993] Combinatorial aspects of convex polytopes, in: *Handbook of Convex Geometry*, eds. P. Gruber and J.M. Wills (North-Holland, Amsterdam) Vol. A, pp. 485–534.

Bing, R.H.

[1964] Some aspects of the topology of 3-manifolds related to the Poincaré conjecture, in: *Lectures on Modern Mathematics*, Vol. II, ed. T.L. Saaty (Wiley, New York) pp. 93–128.

Björner, A.

[1980] Shellable and Cohen–Macaulay partially ordered sets, *Trans. Amer. Math. Soc.* **260**, 159–183.

[1981] Homotopy type of posets and lattice complementation, *J. Combin. Theory A* **30**, 90–100.

[1984a] Posets, regular CW complexes and Bruhat order, *European J. Combin.* **5**, 7–16.

[1984b] Some combinatorial and algebraic properties of Coxeter complexes and Tits buildings, *Adv. in Math.* **52**, 173–212.

[1985] Combinatorics and topology, *Notices Amer. Math. Soc.* **32**, 339–345.

[1992] Homology and shellability of matroids and geometric lattices, in: *Matroid Applications*, ed. N. White (Cambridge University Press, Cambridge), pp. 226–283.

[1994a] Subspace arrangements, in: *Proc. 1st European Congress of Mathematicians, Paris, 1992*, eds. A. Joseph et al. (Birkhäuser, Basel) pp. 321–370.

[1994b] A general homotopy complementation formula, *Discrete Math.*, to appear.

Björner, A., and G. Kalai

[1988] An extended Euler–Poincaré theorem, *Acta Math.* **161**, 279–303.

[1989] On f-vectors and homology, in: *Combinatorial Mathematics – Proc. 3rd Int. Conf., 1985*, eds. G. Bloom et al., *Ann. New York Acad. Sci.* **555**, 63–80.

Björner, A., and M. Wachs

[1983] On lexicographically shellable posets, *Trans. Amer. Math. Soc.* **277**, 323–341.

[1994] Shellable nonpure complexes and posets, *Trans. Amer. Math. Soc.*, to appear.

Björner, A., and J.W. Walker

[1983] A homotopy complementation formula for partially ordered sets, *European J. Combin.* **4**, 11–19.

Björner, A., and G.M. Ziegler

[1992] Introduction to greedoids, in: *Matroid Applications*, ed. N. White (Cambridge University Press, Cambridge) pp. 284–357.

Björner, A., A.M. Garsia and R.P. Stanley

[1982] An introduction to Cohen–Macaulay partially ordered sets, in: *Ordered Sets*, ed. I. Rival (Reidel, Dordrecht) pp. 583–615.

Björner, A., B. Korte and L. Lovász

[1985] Homotopy properties of greedoids, *Adv. in Appl. Math.* **6**, 447–494.

Björner, A., M. Las Vergnas, B. Sturmfels, N.L. White and G.M. Ziegler

[1993] *Oriented Matroids* (Cambridge University Press, Cambridge).

Björner, A., L. Lovász, S.T. Vrećica and R.T. Živaljević

[1994] Chessboard complexes and matching complexes, *J. London Math. Soc. (2)* **49**, 25–39.

Bogatyi, S.A.

[1986] Topological methods in combinatorial problems, *Russian Math. Surveys* **41**, 43–57.

Bokowski, J., and B. Sturmfels

[1989] *Synthetic Computational Geometry, Lecture Notes in Mathematics*, Vol. 1355 (Springer, Berlin).

Borsuk, K.

[1933] Drei Sätze über die n-dimensionale euklidische Sphäre, *Fund. Math.* **20**, 177–190.

[1948] On the imbedding of systems of compacta in simplicial complexes, *Fund. Math.* **35**, 217–234.

Brown, K.S.

[1989] *Buildings* (Springer, Berlin).

Brown, R.

[1968] *Elements of Modern Topology* (McGraw-Hill, London).

Bruggesser, H., and P. Mani

[1971] Shellable decompositions of cells and spheres, *Math. Scand.* **29**, 197–205.

Budach, L., B. Graw, C. Meinel and S. Waack

[1988] *Algebraic and Topological Properties of Finite partially Ordered Sets, Teubner Texte zur Mathematik*, Vol. 109 (Teubner, Leipzig).

Buekenhout, F.

[1995] ed., *Handbook of Incidence Geometry* (North-Holland, Amsterdam).

Cohen, D.I.A.

[1967] On the Sperner lemma, *J. Combin. Theory* **2**, 585–587.

Cohen, M.M.

[1973] *A Course in Simple-Homotopy Type* (Springer, New York).

Cooke, G.E., and R.L. Finney

[1967] *Homology of Cell Complexes* (Princeton University Press, Princeton, NJ).

Csima, J., and E.T. Sawyer

[1993] There exist $6n/13$ ordinary points, *Discrete Comput. Geometry* **9**, 187–202.

Curtis, C.W., G.I. Lehrer and J. Tits

[1980] Spherical buildings and the character of the Steinberg representation, *Invent. Math.* **58**, 201–210.

Danaraj, G., and V. Klee

[1974] Shellings of spheres and polytopes, *Duke Math. J.* **41**, 443–451.

[1978] Which spheres are shellable? *Ann. Discrete Math.* **2**, 33–52.

Daverman, R.J.

[1986] *Decompositions of Manifolds* (Academic Press, New York).

Dowker, C.H.

[1952] Homology groups of relations, *Ann. of Math.* 56, 84–95.

Dugundji, J., and A. Granas

[1982] *Fixed Point Theory,* Vol. I (Polish Scientific Publishers, Warszawa).

Eckhoff, J.

[1979] Radon's theorem revisited, in: *Contributions to Geometry, Proc. Siegen, 1978,* eds. J. Tölke and J.M. Wills (Birkhäuser, Basel) pp. 164–185.

Edelman, P.H.

[1984] The acyclic sets of an oriented matroid, *J. Combin. Theory B* **36**, 26–31.

Edmonds, J., and A. Mandel

[1978] Topology of oriented matroids, *Notices Amer. Math. Soc.* **25**, A-510.

Edmonds, J., L. Lovász and A. Mandel

[1980] Solution, *Math. Intelligencer* **2**, 107.

Ewald, G.

[1995] *Combinatorial Convexity and Algebraic Geometry* (Springer, Berlin).

Folkman, J.

[1966] The homology groups of a lattice, *J. Math. Mech.* **15**, 631–636.

Folkman, J., and J. Lawrence

[1978] Oriented matroids, *J. Combin. Theory B* **25**, 199–236.

Freund, R.M., and M.J. Todd

[1981] A constructive proof of Tucker's combinatorial lemma, *J. Combin. Theory A* **30**, 321–325.

Fulton, W.

[1993] *Introduction to Toric Varieties* (Princeton University Press, Princeton, NJ).

Gale, D.

[1956] Neighboring vertices on a convex polyhedron, in: *Linear Inequalities and Related Systems,* eds. H.W. Kuhn and A.W. Tucker (Princeton University Press, Princeton, NJ) pp. 255–263.

[1979] The game of *HEX* and the Brouwer fixed-point theorem, *Amer. Math. Monthly* **86**, 818–827.

Grünbaum, B.

[1967] *Convex Polytopes* (Interscience–Wiley, London).

Győry, E.

[1978] On division of graphs to connected subgraphs, in: *Combinatorics I,* eds. A. Hajnal and V.T. Sós, *Colloq. Math. Soc. János Bolyai* **18**, 485–494.

Hartshorne, R.

[1977] *Algebraic Geometry* (Springer, Berlin).

Hill, T.P.

[1988] Common hyperplane medians for random vectors, *Amer. Math. Monthly* **95**, 437–441.

Hochster, M.

[1977] Cohen–Macaulay rings, combinatorics and simplicial complexes, in: *Ring Theory II,* eds. B.R. McDonald and R. Morris (Dekker, New York) pp. 171–223.

Hudson, J.F.P.

[1969] *Piecewise Linear Topology* (Benjamin, New York).

Kahn, J., M. Saks and D. Sturtevant

[1984] A topological approach to evasiveness, *Combinatorica* **4**, 297–306.

Korte, B., L. Lovász and R. Schrader

[1991] *Greedoids* (Springer, Berlin).

Las Vergnas, M.

[1978] Bases in oriented matroids, *J. Combin. Theory B* **25**, 283–289.

Lawrence, J.

[1984] *Shellability of Oriented Matroid Complexes*, Preprint.

Leray, J.

[1945] Sur la forme des espaces topologiques et sur les points fixes des représentations, *J. Math. Pures Appl.* **24**, 95–167.

Lickorish, W.B.R.

[1991] Unshellable triangulations of spheres, *European J Combin.* **12**, 527–530.

Lindström, B.

[1981] On matroids and Sperner's lemma, *European J. Combin.* **2**, 65 66.

Lovász, L.

[1977] A homology theory for spanning trees of a graph, *Acta Math. Acad. Sci. Hungar.* **30**, 241 251.

[1978] Kneser's conjecture, chromatic number and homotopy, *J. Combin. Theory A* **25**, 319–324.

[1979] Topological and algebraic methods in graph theory, in: *Graph Theory and Related Topics*, eds. J.A. Bondy and U.S.R. Murty (Academic Press, New York) pp. 1–14.

[1980] Matroids and Sperner's lemma, *European J. Combin.* **1**, 65 66.

[1983] Self-dual polytopes and the chromatic number of distance graphs on the sphere, *Acta Sci. Math. (Szeged)* **45**, 317–323.

Lundell, A.T., and S. Weingram

[1969] *The Topology of CW Complexes* (Van Nostrand, New York).

Mandel, A.

[1982] *Topology of oriented matroids*, Ph.D. Thesis (University of Waterloo).

Mather, J.

[1966] Invariance of the homology of a lattice, *Proc. Amer. Math. Soc.* **17**, 1120–1124.

Maurer, S.B.

[1973] Matroid basis graphs I, *J. Combin. Theory B* **14**, 216–240.

McCord, M.C.

[1966] Singular homology groups and homotopy groups of finite topological spaces, *Duke Math. J.* **33**, 465–474.

[1967] Homotopy type comparison of a space with complexes associated with its open covers, *Proc. Amer. Math. Soc.* **18**, 705–708.

McMullen, P.

[1993] On simple polytopes, *Invent. Math.* **113**, 419–444.

Milnor, J.

[1956] Construction of universal bundles, II, *Ann. of Math.* **63**, 430–436.

Munkres, J.R.

[1984a] *Elements of Algebraic Topology* (Addison-Wesley, Menlo Park, CA).

[1984b] Topological results in combinatorics, *Michigan Math. J.* **31**, 113–128.

Oda, T.

[1988] *Convex Bodies and Algebraic Geometry – An Introduction to the Theory of Toric Varieties* (Springer, Berlin).

Oliver, R.

[1975] Fixed-point sets of group actions on finite acyclic complexes, *Comment. Math. Helv.* **50**, 155–177.

Orlik, P.

[1990] Milnor fiber complexes for Shephard groups, *Adv. in Math.* **83**, 135–154.

Orlik, P., and H. Terao

[1992] *Arrangements of Hyperplanes* (Springer, Berlin).

Pachner, U.

[1986] Konstruktionsmethoden und das kombinatorische Homöomorphieproblem für Triangulationen kompakter semilinearer Mannigfaltigkeiten, *Abh. Math. Sem. Univ. Hamburg* **57**, 69–86.

Proctor, R.

[1982] Representations of $sl(2,\mathbb{C})$ on posets and the Sperner property, *SIAM J. Algebraic Discrete Methods* **3**, 275–280.

Provan, J.S., and L.J. Billera

[1980] Decompositions of simplicial complexes related to diameters of convex polyhedra, *Math. Oper. Res.* **5**, 576–594.

Quillen, D.

[1978] Homotopy properties of the poset of non-trivial p-subgroups of a group, *Adv. in Math.* **28**, 101–128.

Rado, R.

[1946] A theorem on general measure, *J. London Math. Soc.* **21**, 291–300.

Reisner, G.A.

[1976] Cohen–Macaulay quotients of polynomial rings, *Adv. in Math.* **21**, 30–49.

Rival, I.

[1985] The fixed point property, *Order* **2**, 219–221.

Ronan, M.A.

[1981] Coverings of certain finite geometries, in: *Finite Geometries and Designs, London Mathematical Society Lecture Note Series*, Vol. 49, eds. P.J. Cameron et al. (Cambridge University Press, Cambridge) pp. 316–331.

[1989] *Lectures on Buildings* (Academic Press, Orlando, FL).

Rota, G.-C.

[1964] On the foundations of combinatorial theory I. Theory of Möbius functions, *Z. Wahrscheinlichkeitstheorie u. Verw. Gebiete* **2**, 340–368.

Rushing, T.B.

[1973] *Topological Embeddings* (Academic Press, New York).

Sarkaria, K.S.

[1990] A generalized Kneser conjecture, *J. Combin. Theory B* **49**, 236–240.

[1991a] Kuratowski complexes, *Topology* **30**, 67–76.

[1991b] A generalized van Kampen–Flores theorem, *Proc. Amer. Math. Soc.* **111**, 559–565.

[1992] Tverberg's theorem via number fields, *Israel J. Math.* **79**, 317–320.

Schrijver, A.

[1978] Vertex-critical subgraphs of Kneser graphs, *Nieuw Arch. voor Wiskunde (3)* **26**, 454–461.

Smith, S.D.

[1988] Geometric techniques in representation theory, *Geom. Dedicata* **25**, 355–373.

Spanier, E.H.

[1966] *Algebraic Topology* (McGraw-Hill, New York).

Sperner, E.

[1928] Neuer Beweis für die Invarianz der Dimensionszahl und des Gebietes, *Abh. Math. Sem. Univ. Hamburg* **6**, 265–272.

Stanley, R.P.

[1975] The upper bound conjecture and Cohen–Macaulay rings, *Studies in Appl. Math.* **54**, 135–142.

[1977] Cohen–Macaulay complexes, in: *Higher Combinatorics*, ed. M. Aigner (Reidel, Dordrecht) pp. 51–62.

[1979] Balanced Cohen–Macaulay complexes, *Trans. Amer. Math. Soc.* **249**, 139–157.

[1980a] Weyl groups, the hard Lefschetz theorem, and the Sperner property, *SIAM J. Algebraic Discrete Methods* **1**, 168–184.

[1980b] The number of faces of a simplicial convex polytope, *Adv. in Math.* **35**, 236–238.

[1983a] *Combinatorics and Commutative Algebra* (Birkhäuser, Basel).

[1983b] Combinatorial applications of the hard Lefschetz theorem, in: *Proc. Int. Congr. of Mathematicians, Warsaw, 1983* (Polish Scientific Publishers, Warsaw) pp. 447–453.

[1985] The number of faces of simplicial polytopes and spheres, in: *Discrete Geometry and Convexity*, eds. J.E. Goodman et al., *Ann. New York Acad. Sci.* **440**, 212–223.

[1987a] On the number of faces of centrally-symmetric simplicial polytopes, *Graphs Combin.* **3**, 55–66.

[1987b] Generalized *h*-vectors, intersection cohomology of toric varieties, and related results, in: *Commutative Algebra and Combinatorics*, eds. H. Matsumura et al., *Adv. Studies Pure Math.* **11**, 187–213.

[1989] personal communication.

Steinlein, H.

[1985] Borsuk's antipodal theorem and its generalizations and applications: A survey, in: *Méthodes Topologiques en Analyse Non Linéaire, Coll. Sém. de Mathematique Superieure*, Vol. 95, ed. A. Granas (Université de Montréal Press, Montréal) pp. 166–235.

Stong, R.E.

[1966] Finite topological spaces, *Trans. Amer. Math. Soc.* **123**, 325–340.

[1984] Group actions on finite spaces, *Discrete Math.* **49**, 95–100.

Thévenaz, J.

[1987] Permutation representations arising from simplicial complexes, *J. Combin. Theory A* **46**, 121–155.

Tits, J.

[1981] A local approach to buildings, in: *The Geometric Vein (The Coxeter Festschrift)*, eds. C. Davis et al. (Springer, New York) pp. 519–547.

Todd, M.J.

[1976] *The Computation of Fixed Points and Applications, Lecture Notes in Economic and Mathematical Systems*, Vol. 124 (Springer, Berlin).

Tutte, W.T.

[1958] A homotopy theorem for matroids, I, II, *Trans. Amer Math. Soc.* **88**, 144–174.

[1965] Lectures on matroids, *J. Res. Nat. Bur. Standards B* **69B**, 1–47.

[1979] *Selected Papers of W.T. Tutte*, eds. D. McCarthy and R.G. Stanton (Babbage Research Centre, St. Pierre, Manitoba).

Tverberg, H.

[1966] A generalization of Radon's theorem, *J. London Math. Soc.* **41**, 123–128.

Vince, A.

[1985] A non-shellable 3-sphere, *European J. Combin.* **6**, 91–100.

Wachs, M.L., and J.W. Walker

[1986] On geometric semilattices, *Order* **2**, 367–385.

Walker, J.W.

[1981a] *Topology and combinatorics of ordered sets*, Ph.D. Thesis (MIT, Cambridge, MA).

[1981b] Homotopy type and Euler characteristic of partially ordered sets, *European J. Combin.* 2, 373–384.

[1983a] From graphs to ortholattices and equivariant maps, *J. Combin. Theory B* **35**, 171–192.

[1983b] A homology version of the Borsuk–Ulam theorem, *Amer. Math. Monthly* **90**, 466–468.

[1988] Canonical homeomorphisms of posets, *European J. Combin.* **9**, 97–107.

Webb, P.J.

[1987] Subgroup complexes, *Proc. Symp. Pure Math.* 47, 349–365.

Weil, A.

[1952] Sur les théorèmes de de Rham, *Comment. Math. Helv.* **26**, 119–145.

Welker, V.

[1994] Shellability in the lattice of subgroups of a finite group, in: *Proc. Jerusalem Combinatorics Conf 1993*, eds. H. Barcelo and G. Kalai, *Contemporary Math. Series* (AMS, Providence, RI) to appear.

Wu, W.-TS.

[1962] On a theorem of Leray, *Chinese Math.* **2**, 398–410.

Yang, C.-T.

[1955] Continuous functions from spheres to Euclidean spaces, *Ann. of Math.* **62**, 284–292.

Yao, A.C.-C.

[1988] Monotone bipartite graph properties are evasive, *SIAM J. Computing* **17**, 517–520.

Yao, A.C.-C., and F.F. Yao

[1985] A general approach to *d*-dimensional geometric queries, in: *Proc. 17th ACM Symp. on the Theory of Computing* (ACM, New York) pp. 163–168.

Yuzvinsky, S.
[1987] Cohen–Macaulay rings of sections, *Adv. in Math.* **63**, 172–195.

Ziegler, G.M., and R.T. Živaljević
[1993] Homotopy types of subspace arrangements via diagrams of spaces, *Math. Ann.* **295**, 527–548.

Živaljević, R.T., and S.T. Vrećica
[1990] An extension of the ham sandwich theorem, *Bull. London Math. Soc.* **22**, 183–186.
[1992] The colored Tverberg's problem and complexes of injective functions, *J. Combin. Theory A* **61**, 309–318.

Part IV
Applications

CHAPTER 35

Combinatorics in Operations Research

Antoon W.J. KOLEN

Quantitative Economics, University of Limburg, P.O, Box 616, 6200 MD Maastricht, The Netherlands

Jan Karel LENSTRA

Department of Mathematics and Computing Science, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands
and
Centre for Mathematics and Computer Science, Amsterdam, The Netherlands

Contents

HANDBOOK OF COMBINATORICS
Edited by R. Graham, M. Grötschel and L. Lovász

This is a collection of examples of the use of combinatorial techniques in practical decision situations. The emphasis is on the description of real-world problems, the formulation of mathematical models, and the development of algorithms for their solution. We survey related models and applications.

Introduction

The spectacular growth of combinatorics over the past few decades is to some extent due to the diversity and the importance of its applications. Combinatorial problems occur in other branches of mathematics and in computer science, in the natural sciences and in the humanities, and in all kinds of practical decision situations. In addition, the solution of these combinatorial problems has often yielded significant advances and benefits. There is little doubt that its successful use in other fields has greatly stimulated research in the area of combinatorics.

The sample of applications of combinatorics presented in this chapter all arise in situations of planning and design that are usually dealt with in operations research. This discipline is concerned with the investigation of models and methods that support decision making in practice. We do not intend to give a complete survey. Instead, we have tried to select some typical examples with the hope of conveying the flavor of the subject. Our selection process has been guided by the following biases and principles.

First, it has been our purpose to show how to solve real-world problems, not how to construct applications of combinatorial models. With one or two exceptions, each of our examples finds its origin in practice. At the same time, we have preferred those problems that give rise to clean and elegant models, and we have avoided complications that in the present context would only obscure the essence.

Further, most of the problems we discuss occurred in the Netherlands. While this emphasis reflects the limitations of our experience, we do not feel that it has narrowed the scope of our examples.

Finally, we will concentrate on the design and analysis of models and algorithms. Collecting data, writing computer codes, building user interfaces, and getting solutions implemented are equally important stages in the practice of combinatorics. They do not belong, however, to the subject matter of this chapter.

Each of the sections below follows the same outline: we describe a practical problem, formulate one or more mathematical models, present suitable solution methods, and survey related models and applications. It is assumed that the reader is familiar with the fundamentals of combinatorics and combinatorial optimization. We refer, in particular, to chapter 2 on connectivity and network flows, chapter 3 on matchings, chapter 4 on coloring, stable sets and perfect graphs, chapter 28 on optimization, and chapter 30 on polyhedral combinatorics.

1. Traveling salesman

1.1. X-rays and arrays

To demonstrate the versatility of the traveling salesman problem as a model, we will consider two very different problems situations.

The first situation involves the *sequencing of X-ray measurements* in crystallography. One wishes to analyze the detailed structure of a crystal. To this end, the crystal is mounted in a diffractometer, and the intensity of X-rays is measured for a large number of positions of the crystal and the reading device inside the apparatus. Such an experiment may require many thousands of readings. These readings can be made relatively quickly, but the repositioning time between successive measurements is substantial. The readings can be taken in any order, and the question is how one should sequence them so as to minimize the time to complete the experiment.

Bland and Shallcross (1989) encountered problems of this type at the Cornell High Energy Synchrotron Source. For experiments with up to 14 464 readings, they computed sequences for which the total repositioning time is within 1.7% of a lower bound on the optimum. The standard method used for sequencing the measurements produced solutions that are generally between 55% and 90% above the optimum.

The second situation concerns the *clustering of a data array*. Given are two finite sets R and S and a nonnegative matrix $(a_{rs})_{r\in R,s\in S}$, where a_{rs} measures the strength of the relationship between elements $r\in R$ and $s\in S$. One would like to permute the rows and columns of the matrix so as to bring its large elements together. The resulting *clustering* should identify strong relationships between subsets of R and S.

McCormick et al. (1972) argue that clustering a matrix may be useful for problem decomposition and data reorganization. They illustrate this with three examples. The first one arises in *airport design*. R $(=S)$ is a set of 27 facilities that should be available at the airport and that are under the control of the designer; a_{rs} is fixed at 0, 1, 2 or 3 depending on whether facilities r and s have no, a weak, a moderate or a strong interdependence. The permuted matrix should suggest a decomposition of the design problem into subproblems that interact not at all or only in a limited and well-defined way. The second example involves a set R of 53 *aircraft types* and a set S of 37 functions they can perform; $a_{rs}=1$ if aircraft r is suitable for function s, and $a_{rs}=0$ otherwise. The rearranged matrix shows which aircraft are able to perform the same functions and which tasks can be performed by the same aircraft. The third example also deals with an object–attribute array. R is a set of 24 *marketing* techniques, S is a set of 17 marketing applications, $a_{rs}=1$ if technique r has been successfully used for application s, and $a_{rs}=0$ otherwise. Lenstra and Rinnooy Kan (1975) give a fourth example. It deals with an *input–output matrix*. R $(=S)$ is a set of 50 regions on the Indonesian islands, $a_{rs}=1$ if at least 50 tons of rice are annually transported from region r to region s, and $a_{rs}=0$ otherwise.

1.2. Model formulation

Both problems can be modeled as a *traveling salesman problem*. This is the problem of a salesman who, starting from his home city, has to find the shortest tour that takes him exactly once through each of a number of other cities and then back home. Suppose there are n cities and c_{ij} is the distance between cities i and j $(i, j = 1, \ldots, n)$. The salesman is interested in a permutation π of $\{1, \ldots, n\}$ that minimizes.

$$\sum_{i=1}^{n-1} c_{\pi(i)\pi(i+1)} + c_{\pi(n)\pi(1)} \,.$$

Here, $\pi(i)$ is the ith city visited. The traveling salesman problem is *symmetric* if $c_{ij} = c_{ji}$ for all i, j.

It is straightforward to cast the sequencing problem in these terms. We identify the readings with the cities and the repositioning time between two readings with the distance between the corresponding cities. We then add one more city with equal distances to the others, in order to transform the problem of finding an open sequence into that of finding a closed tour. Note that the distances are symmetric.

As to the clustering problem, we first have to convert it into an optimization problem. McCormick et al. (1972) propose to measure the effectiveness of a clustering by the sum of all products of horizontally or vertically adjacent elements. The reader can easily convince himself that higher sums of these products tend to correspond to better clusterings. The problem is now to permute the rows and columns of the matrix so as to maximize this criterion.

Permuting the rows does not affect the horizontal adjacencies of the elements, and permuting the columns does not affect their vertical adjacencies. The problem therefore decomposes into two separate and similar problems, one for the rows and one for the columns. We consider the former. The row optimization problem is to find a permutation ρ of R that maximizes

$$\sum_{r=1}^{|R|-1} \sum_{s \in S} a_{\rho(r)s} a_{\rho(r+1)s} \,.$$

Here, row $\rho(r)$ of the matrix is put in position r. This is, again, nothing but the symmetric traveling salesman problem in disguise (Lenstra 1974). Let $R = \{1, \ldots, |R|\}$, and define

$$n = |R| + 1 \,,$$

$$c_{ij} = -\sum_{s \in S} a_{is} a_{js} \,, \qquad c_{in} = c_{nj} = 0 \qquad \text{for } i, j \in R \,.$$

The rows of the matrix are the cities, the additive inverses of their inner products are the distances, and a dummy city has been added to close the tour.

The symmetric traveling salesman problem is conveniently formulated in terms of undirected graphs. Consider the complete graph $K_n = (V, E)$ on n nodes, where a weight c_e is associated with each edge $e \in E$. The problem is to find a *Hamiltonian circuit* or *tour* in K_n, i.e., a circuit that visits each node exactly once, of minimum total weight. This generalizes the problem of determining whether a given graph contains a Hamiltonian circuit, which is NP-complete.

An *integer programming* formulation is as follows. Let $x_e = 1$ indicate that the salesman travels along the edge e. The problem is then to minimize

$$\sum_{e \in E} c_e x_e$$

subject to

$$\sum_{e \in \delta(i)} x_e = 2 \quad \text{for all } i \in V , \tag{1.1}$$

$$\sum_{e \in \delta(U)} x_e \geq 2 \quad \text{for all } U \subset V \text{ , } U \neq \emptyset \text{ , } U \neq V \text{ ,} \tag{1.2}$$

$$x_e \in \{0, 1\} \quad \text{for all } e \in E \text{ .} \tag{1.3}$$

Here, $\delta(U)$ is the set of edges with exactly one end in U; we write $\delta(i)$ for $\delta(\{i\})$. Without the constraints (1.2), this is the *2-matching problem*; each node has degree 2, but the selected edges do not necessarily form a single tour. The constraints (1.2) eliminate subtours on any proper subset $U \subset V$.

1.3. Solution approaches

Many types of *approximation algorithms* have been developed for the symmetric traveling salesman problem. It is useful to distinguish between *constructive* methods, which build a single tour, and *iterative improvement* methods, which search the neighborhood of the current solution for a better one and continue the search until a local optimum has been obtained. An example of a constructive method is the *nearest neighbor rule*: the salesman starts in a given city and always travels to an unvisited city that is closest to the last chosen city. One of the earliest iterative improvement methods was proposed by Lin (1965). He defines the *k-exchange neighborhood* of a tour as the set of all tours that can be obtained from it by replacing any set of k edges by another set of k edges, and he calls a tour that is locally optimal with respect to this neighborhood *k-opt*. The values $k = 2$ and $k = 3$ are most often used. The champion among heuristics for the symmetric traveling salesman problem is the *variable-depth search* method of Lin and Kernighan (1973), where the value of k is not specified in advance.

In order to enhance the computational efficiency of edge exchange procedures on large problem instances, one usually replaces K_n by a sparse subgraph. Bland and Shallcross (1989), for example, included for each node only the ten shortest incident edges. Their upper bounds were computed by the Lin–Kernighan algorithm, and their lower bounds by the Held–Karp algorithm mentioned below.

Optimization algorithms for the symmetric traveling salesman problem usually proceed by branch and bound, with lower bounds based on spanning 1-trees combined with Lagrangean relaxation, or on fractional 2-matchings combined with cutting planes.

A *spanning* 1-*tree* consists of a spanning tree on the node set $V\setminus\{1\}$ and two edges incident to node 1. A minimum-weight spanning 1-tree can be found in polynomial time. In comparison with a tour, it is still a connected graph with n edges, but the requirement that all node degrees should be equal to 2 has been relaxed. Its weight is therefore a lower bound on the length of a shortest tour. The lower bound may be improved by calculating a penalty d_i for each node i (positive if the degree of i is larger than 2, negative if it is smaller) and to replace the weight c_e by $c_e + d_i + d_j$ for each edge $e = \{i, j\}$. The ordering of tours according to length is invariant under this transformation, but the optimal spanning 1-tree may change. The approach is due to Held and Karp (1970, 1971) and signified the beginning of the use of Lagrangean relaxation in combinatorial optimization. The penalties or *Lagrangean multipliers* d_i are calculated by general subgradient optimization techniques or by special multiplier adjustment schemes.

A *fractional* 2-*matching* is a feasible solution to (1.1) and $0 \leqslant x_e \leqslant 1$ $(e \in E)$. In comparison with a tour, the subtour elimination constraints (1.2) and the integrality requirements in (1.3) have been relaxed. An optimal fractional 2-matching can be obtained in polynomial time, e.g., by linear programming. The resulting lower bound can be improved by the addition of *cutting planes* that correspond to facets of the symmetric traveling salesman polytope, i.e., the convex hull of all solutions to (1.1)–(1.3). The subtour elimination constraints (1.2) define facets, but many more classes of facets have been identified. Given the solution corresponding to such a lower bound, one tries to find a violated facet and adds the corresponding constraint to the linear program. If the sequence of linear programs does not yield a feasible solution to (1.1)–(1.3) (and hence an optimal tour), then some form of tree search is applied. There are, in general, two difficulties with this *polyhedral approach*. First, it is very unlikely that one will ever be able to characterize all of the facets. Secondly, the so-called *separation problem* of finding a facet that is violated by the solution to the linear program is often far from trivial; usually, fast heuristics are used. However, Padberg and Rinaldi (1987) and Grötschel and Holland (1991) have obtained impressive computational results with this approach.

For more detailed information on branch and bound, Lagrangean relaxation, polyhedral techniques, and the relation between the optimization problem and the separation problem, see chapters 28 and 30.

1.4. Related models and applications

As has already been observed, the problem of determining whether a given graph contains a Hamiltonian circuit is a special case of the traveling salesman problem. This, in turn, generalizes to the *vehicle routing problem*, which is the subject of section 2.

A quite different class of routing problems emerges if one wishes to visit all of the *edges* (or *arcs*) of a graph rather than all of the nodes. The basic problem is then to determine if a given graph contains an *Eulerian tour*, i.e., a closed walk that traverses each edge (or arc) exactly once. This generalizes to the *Chinese postman problem*, where one has to find the shortest closed walk that traverses each edge (or arc) at least once. It will arise in the solution of a practical multi-postman problem in section 3.

The traveling salesman problem occurs in many more practical situations than sequencing measurements, clustering matrices, or routing vehicles. Ratliff and Rosenthal (1983) describe the problem of order picking in a rectangular warehouse. This is a traveling salesman problem that, due to the structure of the underlying network, can be solved in polynomial time by dynamic programming techniques. Other applications are discussed by Lenstra and Rinnooy Kan (1975).

The traveling salesman problem has become the prototypical problem of combinatorial optimization. This is partly because its simplicity of statement and difficulty of solution are even more apparent than for most other problems in the area. In addition, many of the solution approaches that have become standard in combinatorial optimization were first developed and tested in the context of the traveling salesman problem. Our presentation of mathematical formulations, solution approaches and applications is only meant to be illustrative. A full treatment of the problem justifies a book of its own (Lawler et al. 1985).

2. Vehicle routing

2.1. *CAR*

In the period 1983–1986, the Centre for Mathematics and Computer Science (CWI) in Amsterdam was involved in the development of a computer system for vehicle routing. The resulting system is called CAR, which stands for 'Computer Aided Routing'. Before we discuss the models and the algorithms that form the mathematical basis of CAR, we review some aspects of the practical background and the computer implementation in this section. The reader is referred to Savelsbergh (1992) for details.

CAR has been designed for the solution of the *single-depot vehicle routing problem with time windows*. A problem situation of this type that occurred at the hanging garment division of a Dutch road transportation firm was our main source of information and motivation from practice. The situation is basically as follows. About fifteen vehicles are stationed at a single central depot and must serve about 500 geographically dispersed customers. Each vehicle has as given capacity. Each customer has a given demand and must be served within a specified time interval. The travel times between the locations of the depot and the customers are given. We have to find a collection of routes for the vehicles, each starting and finishing at the depot and collectively visiting all customers, while respecting the capacity constraints of the vehicles and the time constraints of the customers. We would like to minimize the total travel time.

Problems of this type and size must be solved daily. There are several reasons why, at the present time, it is not possible to completely automate their solution.

On the one hand, the models that arise are *hard*, in a well-defined sense. Better solutions are obtained if the computer system and its user cooperate and divide the tasks in accordance with their respective – and complementary – capabilities. On the other hand, the problem situations are *soft*. Feasibility constraints can be stretched, and optimality on one criterion will be weighed against the values of secondary criteria. This is not the place to advocate the benefits of man–machine interaction in complex decision situations. Suffice it to refer to Anthonisse et al. (1988) and to mention that CAR has been designed and built as an interactive system, which does not make decisions but only supports decision making by the people who are in charge. The system is being used by several firms in the Netherlands.

2.2. Model formulation

We present an integer programming formulation of the single-depot vehicle routing problem. For the time being, we ignore the time windows of the customers.

The data of the problem are as follows. There are m vehicles. The capacity of vehicle h is equal to Q_h $(h = 1, \ldots, m)$. The depot is indexed by $i = 1$ and the customers by $i = 2, \ldots, n$; the demand of customer i is equal to q_i $(i = 2, \ldots, n)$. Finally, there is a matrix $(c_{ij})^n_{i,j=1}$ of travel times. As to the decision variables, let

$$y_{hi} = \begin{cases} 1 & \text{if vehicle } h \text{ visits customer } i\text{ ,} \\ 0 & \text{otherwise ,} \end{cases}$$

$$x_{hij} = \begin{cases} 1 & \text{if vehicle } h \text{ visits customers } i \text{ and } j \text{ in sequence ,} \\ 0 & \text{otherwise .} \end{cases}$$

The problem is now to minimize

$$\sum_{h=1}^{m} \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{hij}$$

subject to

$$\sum_{h=1}^{m} y_{hi} = \begin{cases} m & \text{for } i = 1\text{ ,} \\ 1 & \text{for } i = 2, \ldots, n\text{ ,} \end{cases} \tag{2.1}$$

$$\sum_{i=2}^{n} q_i y_{hi} \leqslant Q_h \qquad \text{for } h = 1, \ldots, m\text{ ,} \tag{2.2}$$

$$y_{hi} \in \{0, 1\} \qquad \text{for } h = 1, \ldots, m\text{ , } i = 1, \ldots, n\text{ ,} \tag{2.3}$$

$$\sum_{j=1}^{n} x_{hij} = \sum_{j=1}^{n} x_{hji} = y_{hi} \quad \text{for } h = 1, \ldots, m\text{ , } i = 1, \ldots, n\text{ ,} \tag{2.4}$$

$$\sum_{i,j\in U} x_{hij} \leq |U| - 1 \qquad \text{for } h = 1, \ldots, m\,,\ U \subset \{2, \ldots, n\}\,, \tag{2.5}$$

$$x_{hij} \in \{0, 1\} \qquad \text{for } h = 1, \ldots, m\,,\ i, j = 1, \ldots, n\,. \tag{2.6}$$

The conditions (2.1) ensure that each customer is allocated to one vehicle and that the depot is allocated to each vehicle. The conditions (2.2) are the vehicle capacity constraints. The conditions (2.4) ensure that a vehicle which arrives at a customer also leaves that customer. The conditions (2.5) are the subtour elimination constraints, in a form that differs from (1.2).

This formulation is due to Fisher and Jaikumar (1981). They observed that it consists of a number of interlinked subproblems, namely, a generalized assignment problem and a collection of m traveling salesman problems. The *generalized assignment problem* is the problem of minimizing.

$$\sum_{h=1}^{m} f_h(y_{h1}, \ldots, y_{hn})$$

subject to (2.1)–(2.3). Here, the y_{hi} are the decision variables and $f_h(y_{h1}, \ldots, y_{hn})$ is the minimum time duration of a tour through the depot and the cluster of customers defined by $\{i \mid y_{hi} = 1\}$. These are *traveling salesman problems*, i.e., $f_h(y_{h1}, \ldots, y_{hn})$ is to the minimum value of

$$\sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{hij}$$

subject to (2.4)–(2.6). Here, the x_{hij} are the decision variables and the y_{hi} prescribe the allocation of customers to vehicles.

The traveling salesman problem is NP-hard, and so is the generalized assignment problem, even if its criterion function is linear in the y_{hi}. However, these subproblems have been well studied.

2.3. Solution approaches

Fisher and Jaikumar originally proposed to solve the single-depot vehicle routing problem to optimality by an iterative process, which can be viewed as an application of Benders decomposition. Replacing each $f_h(y_{h1}, \ldots, y_{hn})$ by a lower linear support, they solve the generalized assignment problem, which provides a *lower bound* on the overall solution value and a tentative *clustering* of the customers into vehicles. They then solve the m resulting traveling salesman problems, which yields an *upper bound* on the overall solution value and a tentative *routing* for each vehicle. In the second iteration, the generalized assignment problem is solved again, with an improved lower linear support derived from the solution obtained in the first iteration, and the process continues. As soon as lower bound and upper bound are equal, an optimal solution has been obtained.

This approach is notable for its conceptual value, not for its computational efficiency. It motivated the development of an approximation algorithm. This is a *cluster first-route second* approach, which essentially consists of the first iteration of the optimization procedure. Much depends on the linearization of the functions f_h that is chosen. Fisher and Jaikumar (1981) propose to select a *seed point* $s(h)$ for each vehicle h; $s(h)$ is a customer who is centrally located in the area that is to be covered by vehicle h. They compute the "extra mileage costs" $d_{hi} = c_{1i} + c_{is(h)} - c_{1s(h)}$, which should approximate the routing costs incurred if customer i is served by vehicle h. They then replace each $f_h(y_{h1}, \ldots, y_{hn})$ by $\sum_{i=1}^{n} d_{hi} y_{hi}$ and solve the linear generalized assignment problem. Finally, they solve a traveling salesman problem for each collection of customers allocated to the same vehicle.

Large generalized assignment problems can be solved close to optimality (Fisher et al. 1986). Large traveling salesman problems are usually solved by edge exchange methods of the type discussed in section 1.3.

We should draw the reader's attention to a small but crucial problem that we encountered during the development of CAR. For an unconstrained traveling salesman problem, it takes constant time to process a single edge exchange, as long as the number of edges involved is bounded by a constant. In the presence of time windows, however, testing feasibility of the route that results from an edge exchange requires an amount of time that is linear in the number of cities. Savelsbergh (1990) developed techniques for implementing local search subject to time windows without an increase in overall time complexity. He extended these techniques to handle other side constraints such as multiple time windows per customer, mixed collections and deliveries, and precedence constraints.

2.4. *Related models and applications*

Vehicle routing problems can be modeled and solved in many different ways. Surveys are given by Bodin et al. (1983) and Christofides (1985).

Desrochers et al. (1988) review the state of the art regarding routing with time window constraints. Next to standard vehicle routing problems with time windows, which are especially relevant in the context of school bus routing, they discuss pickup and delivery problems with time windows, which arise in dial-a-ride situations. The characteristic difference between the two problem types is that, in the latter case, pickup and delivery of the same commodity occurs in a single route. The models they present are based on integer programming, dynamic programming, and set partitioning.

After this discussion of node routing problems, the next section deals with *arc routing*. The salesman is replaced by the postman.

3. Multiple postmen

3.1. *Sprinkling highways*

In the winter, the highways in the Netherlands are sprinkled with salt to prevent them from becoming slippery. Some highways have built-in sensors, which indicate when the road temperature drops below a certain threshold; in other cases, the critical point in time is determined by visual inspection. Safety regulations require that, when the signal for preventive sprinkling is given, all highways in a region should be handled within a time period of 45 minutes. The salt sprinklers are stationed at various depots along the highways. They can carry an amount of salt that suffices for a period of 45 minutes. When sprinkling, they drive at a reduced speed. The highway system is such that a road segment may have to be traversed without being sprinkled. How many sprinklers are needed, and how should their routes be constructed?

This is a slight simplification of a problem that was handled by ORTEC Consultants in Gouda, in cooperation with the first author. The project resulted in a prototype program that, on a small problem instance, reduced the number of routes from seventeen to thirteen. Actual instances involve about 150 depots and 500 routes.

3.2. Model formulation

Highways are one-way streets. The highway system is therefore modeled as a directed graph. The road junctions and the salt depots are the nodes, and the road segments are the arcs. The graph is strongly connected. Each arc has two weights, indicating the time needed to traverse the arc while sprinkling and while driving without sprinkling, respectively. A feasible solution is a collection of directed walks, one for each vehicle, such that each walk starts at one of the depot nodes, it is indicated for each occurrence of an arc in a walk whether it is sprinkled or not, each arc is sprinkled exactly once, and no walk exceeds a given upper bound in length. Note that a walk does not have to be closed; the time the vehicles need to return to their depots is irrelevant. A solution is optimal if the sum of the squared differences between upper bound and actual walk lengths is minimum. This criterion models the objective to make the walk lengths large on the one hand and more or less equal on the other.

The sprinkling problem belongs to the class of *arc routing* problems, which was already mentioned in section 1.4. In designing a solution method for our problem, we will relate it to the standard arc routing problem, the *directed Chinese postman problem*. This problem is formulated as follows: given a strongly connected arc-weighted directed graph, find the shortest closed directed walk that traverses each arc at least once. It is solvable in polynomial time, as will be indicated in section 3.3.

Our problem is essentially a *multi-postman problem* with some complexifying characteristics: there are several depots, to which the postmen do not have to return, and there is a quadratic cost function. Irrespective of the objective, the problem of deciding if two postmen stationed at one depot can do the job is already NP-complete.

3.3. Solution approach

We follow an approximative solution strategy. In contrast to the approach of section 2.3, it is a *route first-cluster second* algorithm. That is, we first construct a single closed directed walk that contains all the arcs and has minimum length, and we then break it up into smaller directed walks, one for each vehicle. The routing problem is nothing but the directed Chinese postman problem; the clustering problem will be solved by dynamic programming.

We first discuss the routing phase. The closed walk that is to be determined may have to traverse some of the arcs more than once. Suppose that, if it traverses an arc k times, we add $k-1$ duplicate arcs to the graph. The closed walk is thereby transformed into an *Eulerian tour*, which traverses each are exactly once. Recall that a directed graph is Eulerian (i.e. has an Eulerian tour) if and only if it is strongly connected and, for each node, the indegree is equal to the outdegree. The directed Chinese postman problem is therefore equivalent to finding a minimum-weight collection of duplicate arcs, the addition of which makes the indegree and outdegree of each node equal to each other.

Let V^+ be the set of nodes for which the indegree is larger than the outdegree; let d_i^+ denote the difference for each $i \in V^+$. Let V^- be the set of nodes for which the outdegree is larger than the indegree; let d_j^- denote the difference for each $j \in V^-$. Note that $\sum_{i\in V^+} d_i^+ = \sum_{j \in V^-} d_j^-$. Further, let c_{ij} be equal to the length of the shortest path from $i \in V^+$ to $j \in V^-$. The decision variables x_{ij} will indicate the number of duplicates of the shortest path from $i \in V^+$ to $j \in V^-$ that have to be added to the graph. The problem is to find x_{ij} for which

$$\sum_{i\in V^+}\sum_{j\in V^-} c_{ij}x_{ij}$$

is minimized subject to

$$\sum_{i\in V^+} x_{ij} = d_j^- \quad \text{for all } j \in V^+ ,$$

$$\sum_{j\in V^-} x_{ij} = d_i^+ \quad \text{for all } i \in V^+ ,$$

$$x_{ij} \in \mathbb{N} \cup \{0\} \quad \text{for all } i \in V^+,\ j \in V^- .$$

This is the *linear transportation problem*, which can be solved in polynomial time (see chapter 2).

In the sprinkling problem, each arc has to be sprinkled only once, so that we have to compute the shortest path lengths c_{ij} using the arc weights that correspond to driving without sprinkling. We solve the linear transportation problem and add duplicate arcs to the graph in accordance with its optimal solution. The resulting graph usually contains many Eulerian tours. We select one using a heuristic rule, which incorporates various secondary criteria that are beyond the scope of this discussion.

We now turn to the clustering phase. We choose a starting point of the Eulerian tour and list the arcs in the order in which they occur, say, $a_1, a_2, \ldots, a_m$. Each arc a_h has a weight w_h. If there are several occurrences of the same arc, then the first one is assumed to be the original arc (and has the higher weight) and the others are the duplicates (with the lower weight). We will restrict our attention to walks that correspond to subsequences of $(a_1, a_2, \ldots, a_m)$.

Let W denote the given upper bound on walk length. For each arc a_h ($h = 1, \ldots, m$), let v_h be the shortest distance from any depot to the tail of a_h, and let I_h be the set of indices i such that the subsequence $(a_h, \ldots, a_i)$ can be traversed by a single vehicle, i.e., $v_h + w_h + \cdots + w_i \leq W$ for all $i \in I_h$. The minimum cost z_h of an optimal partitioning of the subsequence $(a_h, \ldots, a_m)$ into feasible walks can now be computed by a simple recursion:

$$z_{m+1} = 0 ,$$

$$z_h = \min_{i \in I_h} \{(W - (v_h + w_h + \cdots + w_i))^2 + z_{i+1}\} \quad \text{for } h = m, m-1, \ldots, 1 .$$

The optimum solution has value z_1. Since walks starting with duplicate arcs may be disregarded, we can restrict the computation to those indices h for which a_h is an original arc, and define $z_h = z_{h+1}$ if a_h is a duplicate arc.

Although the directed Chinese postman problem in the routing phase and the partitioning problem in the clustering phase are both solved to optimality, the solution obtained is only approximate. This is due to the decomposition of the solution process into two phases and to the heuristic choice of an Eulerian tour. The entire algorithm runs in polynomial time.

3.4. *Related models and applications*

The Chinese postman problem was originally formulated on *undirected* graphs, by Guan (1962). It can be solved by shortest path and matching techniques; see, e.g., Lawler (1976). While both the undirected and the directed case can be solved in polynomial time, the postman problem becomes NP-hard if it is *mixed*, *windy*, *rural*, *multiple*, or *capacitated*, and also if the postman is replaced by a *stacker crane*. Lenstra and Rinnooy Kan (1981) and Johnson and Papadimitriou (1985) review complexity results and approximation algorithms for these variants.

Shortest path algorithms are discussed by Lawler (1976). They are used as subroutines in many other combinatorial algorithms, e.g., for the linear transportation problem and the minimum cost flow problem (see chapter 2).

Knuth and Plass (1981) describe an interesting applications of shortest paths that arose during the development of the TEX text processing system. A paragraph of text is to be broken into lines. Nodes correspond to feasible breaking points, and arcs to feasible lines. With each arc, a weight is associated that measures the quality of the breaks at its endpoints and of the line in between. The determination of these weights is not easy, but it is a typographical rather than a mathematical question. As the resulting directed graph is acyclic, a shortest path can be found by a very simple algorithm.

4. Linear ordering

4.1. *Ranking priorities*

In 1970, a Dutch trade union was planning its policy for the future. Nine action items were listed:

(1) increasing the retirement payments as well as the pensions for widows and orphans;

(2) increasing the payment for loss of working hours due to frost;

(3) increasing the holiday allowance;

(4) increasing the pensions for widows and orphans;

(5) introduction of capital growth sharing;

(6) reduction of working hours;

(7) increasing the number of days off;

(8) education of young people at full pay;

(9) increasing the retirement payments.

Being a democratic organization, the trade union decided to involve its members. One thousand union representatives and 326 ordinary members were asked to rank the items in order of decreasing importance. But how does one aggregate 1 000 or 326 individual rankings into a single one? Anthonisse (private communication) proposed a simple and elegant model.

4.2. Model formulation

By ranking n items, an individual expresses $n(n-1)/2$ preferences, one for each pair of items. One way to evaluate an overall ranking is by counting the total number of individual preferences that are consistent with it.

Suppose c_{ij} is the number of people who prefer item i to j, for $i, j = 1, \ldots, n$. An ordering ρ of $\{1, \ldots, n\}$ defines a priority ranking in the sense that i is ranked higher than j if $\rho(i) < \rho(j)$. The total number of preferences that are consistent with a ranking ρ is given by

$$\sum_{i,j:\ \rho(i)<\rho(j)} c_{ij} ,$$

and the problem is to find a ranking ρ that maximizes this number. This is the *linear ordering problem*. It generalizes the *feedback arc set problem* (see section 4.4) and is therefore NP-hard.

A formulation in terms of 0–1 variables is easily obtained. Let $x_{ij} = 1$ indicate that $\rho(i) < \rho(j)$. The problem is then to maximize

$$\sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{ij}$$

subject to

$$x_{ii} = 0 \qquad \text{for } i = 1, \ldots, n , \tag{4.1}$$

$$x_{ij} + x_{ji} = 1 \qquad \text{for } i, j = 1, \ldots, n ,\ i < j , \tag{4.2}$$

$$x_{ij} + x_{jk} + x_{ki} \le 2 \qquad \text{for } i, j, k = 1, \ldots, n ,\ i < j < k \text{ or } i < k < j , \tag{4.3}$$

$$x_{ij} \in \{0, 1\} \qquad \text{for } i, j = 1, \ldots, n . \tag{4.4}$$

The conditions (4.3) represent the transitivity of ρ: if we rank i above j and j above k ($x_{ij} = x_{jk} = 1$), then we rank i above k ($x_{ki} = 0$, so $x_{ik} = 1$).

4.3. *Solution approaches*

For the trade union's problem, Anthonisse replaced (4.4) by $x_{ij} \in \{0, 1\}$ by $x_{ij} \geqslant 0$, solved the resulting linear programming problem, and obtained an integral solution. It has been observed more often that, for this problem type, the linear programming relaxation gives an optimal solution to the integer program or at least an excellent upper bound.

The best available algorithm for the linear ordering problem uses polyhedral techniques, very much in the spirit of the polyhedral approach to the symmetric traveling salesman problem (see section 1.3). The constraints (4.3) define facets of the linear ordering polytope, but, again, more classes of facets have been identified. Reinelt (1985) describes this approach in detail. He also reviews earlier optimization and approximation algorithms for the linear ordering problem.

4.4. *Related models and applications*

Another application of the linear ordering problem is the triangulation of input–output matrices. Here, there are n industry sectors and c_{ij} denotes the supply from sector i to sector j. The sectors have to be ordered "from raw material to consumer".

The linear ordering problem is equivalent to the *acyclic subgraph problem*: given a directed graph $G = (V, A)$ with weights associated with the arcs, find an acyclic directed graph $G' = (V, A')$ with $A' \subset A$ such that the sum of the weights of the arcs in A' is maximum. If all arc weights are equal, the problem reduces to the *feedback arc set problem*: given a directed graph $G = (V, A)$, find a minimum-cardinality set of arcs that intersects each directed circuit in G. We leave it to the reader to sort out the details and refer to Jünger (1985) and Reinelt (1985) for further information on models, algorithms and applications.

5. Clique partitioning

5.1. *Distinguishing types of professions*

In 1969, the Interfaculty of Actuarial Sciences and Econometrics of the University of Amsterdam wished to revise the curriculum in econometrics. "Econometrics" is used here in a broad sense and includes mathematical economics, empirical econometrics, statistics, and operations research. A committee was installed to investigate what professional econometricians in practice would demand of a curriculum.

The committee interviewed 45 econometricians employed in government, industry and consultancy. Each of them was given a list of 24 problem situations and activities and a list of 24 methods and techniques, and was asked to indicate which problems he had been working on during the last two years and which techniques he had applied. The committee then first analyzed the lists of problems to determine which types of professionals could be distinguished. Secondly, the lists of techniques were used to design a curriculum for each type. It should be noted that no one intended to directly implement the results. There are many reasons why an actual university curriculum could differ from what present practice views as desirable.

The first problem was modeled and solved by techniques from combinatorial optimization; details will be given in sections 5.2 and 5.3. The result was a clear distinction between two groups: the "measurers" (proper econometricians) and the "regulators" (operations researchers). The second problem was a statistical exercise, which need not concern us here. A full account is given by Cramer et al. (1970).

5.2. Model formulation

Given are a set V of persons and a set P of problem situations. For each person $i \in V$ there is a set $P_i \subset P$ of problems on which i has worked. For each problem $p \in P$, there is a set $V_p \subset V$ of persons who worked on p, with $V_p = \{i : p \in P_i\}$. We will write $V = \{1, \ldots, n\}$, $\bar{P}_i = P \backslash P_i$ $(i \in V)$, and $\bar{V}_p = V \backslash V_p$ $(p \in P)$.

We wish to partition V into a number of mutually disjoint groups in such a way that two persons i and j are allocated to the same group if P_i and P_j are similar and to different groups if P_i and P_j are dissimilar. In other words, the partition X of V we look for should be a reasonable aggregation of the given partitions $(V_p, \bar{V}_p)$ $(p \in P)$. In the case that $|P| = 1$, we could take $X = (V_1, \bar{V}_1)$, and there would be complete agreement between input and output. In the general case, we choose to minimize the sum, over all problems $p \in P$, of the number of disagreements between X and $(V_p, \bar{V}_p)$.

The data can be represented by numbers y_{ijp} $(1 \leq i < j \leq n,\ p \in P)$ such that $y_{ijp} = 1$ if i and j are in agreement regarding problem p, i.e., $p \in (P_i \cap P_j) \cup (\bar{P}_i \cap \bar{P}_j)$, and $y_{ijp} = 0$ otherwise. Similarly, X can be described by decision variables x_{ij} $(1 \leq i < j \leq n)$ such that $x_{ij} = 1$ if i and j are allocated to the same group and $x_{ij} = 0$ otherwise. Since $z^2 = z$ for any $z \in \{0, 1\}$, the total number of disagreements can now be written as

$$\begin{aligned}
&\sum_{p \in P} \sum_{1 \leq i < j \leq n} (x_{ij} - y_{ijp})^2 \\
&\quad = \sum_{p} \sum_{i<j} (1 - 2y_{ijp}) x_{ij} + \sum_{p} \sum_{i<j} y_{ijp} \\
&\quad = \sum_{i<j} c_{ij} x_{ij} + C ,
\end{aligned}$$

where

$$c_{ij} = \sum_{p} (1 - 2y_{ijp}) = |P| - 2(|P_i \cap P_j| + |\bar{P}_i \cap \bar{P}_j|) ,$$

$$C = \sum_{p} \sum_{i<j} y_{ijp} = \sum_{i<j} (|P_i \cap P_j| + |\bar{P}_i \cap \bar{P}_j|) .$$

constant term C can be dropped. The problem is then to minimize

$$\sum_{1 \leq i < j \leq n} c_{ij} x_{ij}$$

subject to the condition that the x_{ij} define a partition of V:

$$\begin{aligned}
&x_{ij} + x_{jk} - x_{ik} \leq 1 && \text{for } i, j, k = 1, \ldots, n \,,\ i < j < k \,, \\
&x_{ij} - x_{jk} + x_{ik} \leq 1 && \text{for } i, j, k = 1, \ldots, n \,,\ i < j < k \,, \\
&-x_{ij} + x_{jk} + x_{ik} \leq 1 && \text{for } i, j, k = 1, \ldots, n \,,\ i < j < k \,, \\
&x_{ij} \in \{0, 1\} && \text{for } i, j = 1, \ldots, n \,,\ i < j \,.
\end{aligned}$$

Note that the problem has a trivial solution if all c_{ij} have the same sign.

The problem can also be formulated in terms of graphs. Consider the complete graph $K_n = (V, E)$ on n nodes with a weight $c_e \in \mathbb{Z}$ for each edge $e \in E$. The problem is to partition V into non-overlapping subsets so as to minimize the sum of the weights of the edges whose ends are in the same subset. This is called the *clique partitioning problem*, and it is known to be NP-hard.

The above discussion follows Grötschel and Wakabayashi (1989). Note that the number of groups in the partition is not specified in advance but computed as part of the solution.

Let us briefly consider the case in which the number of groups is not free but fixed to, say, m. This *clique m-partitioning problem* can be stated as follows: given the complete graph $K_n = (V, E)$ on n nodes with a weight $c_e \in \mathbb{Z}$ for each edge $e \in E$, color each node with one of m colors so as to minimize the sum of the weights of the edges whose ends receive the same color. This formulation generalizes the *graph coloring problem*, where all c_e are equal to 0 or 1 and we are interested in the existence of an m-coloring with value 0. The graph coloring problem is solvable in polynomial time for $m = 2$ and NP-complete for any $m \geq 3$. The clique 2-partitioning problem is also known as the *max cut problem*, which is already NP-hard in itself.

Carlson and Nemhauser (1966) were the first to consider the clique m-partitioning problem. They gave a *quadratic programming* formulation. Let $x_{hi} = 1$ indicate that node i receives color h. The problem is then to minimize

$$\tfrac{1}{2} \sum_{h=1}^{m} \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} x_{hi} x_{hj}$$

subject to

$$\begin{aligned}
&\sum_{h=1}^{m} x_{hi} = 1 && \text{for } i = 1, \ldots, n \,, \\
&x_{hi} \in \{0, 1\} && \text{for } h = 1, \ldots, m \,,\ i = 1, \ldots, n \,.
\end{aligned}$$

It is not hard to see that, if $x_{hi} \in \{0, 1\}$ is replaced by $x_{hi} \geq 0$, then there exists an integral optimal solution. It can also be proved that an integral feasible solution satisfies the Karush–Kuhn–Tucker conditions if and only if it cannot be improved by giving any single node another color.

5.3. Solution approaches

Grötschel and Wakabayashi (1989) developed a cutting plane algorithm for the clique partitioning problem, along the same lines as the polyhedral approaches mentioned in sections 1.3 and 4.3. They were able to solve problem instances with up to 158 nodes, without ever having to resort to tree search.

In 1969, when the problem at the University of Amsterdam occurred, a more heuristic approach was taken. First, for various values of m, the number of groups was fixed at m and a good partitioning into m groups was computed. Secondly, the results were analyzed and an appropriate value of m was determined. The weights were defined by

$$c_{ij} = \frac{|P| - |P_i \cap P_j| - 0.9|\bar{P}_i \cap \bar{P}_j|}{|P_i \cap P_j| + 1} .$$

Note that a positive agreement between i and j counts more heavily than a negative agreement. This choice was made after some experiments with small problem instances. The results obtained were, however, quite robust with respect to small perturbations of the weights.

None of the optimization algorithms for the clique m-partitioning problem that were available in 1969 could handle instances with $n = 45$. An iterative improvement procedure was developed, with k $(1 \leq k \leq n)$ as an input parameter. At each step, the existing partitioning is considered and the k persons are determined whose individual transition to another group would result in the largest decrease of the criterion value. These k persons are then optimally reallocated by a branch and bound algorithm. If no further improvements are possible, a local optimum has been obtained, which could be called *k-opt*. As mentioned before, a solution is 1-opt if and only if it satisfies the Karush–Kuhn–Tucker conditions of the Carlson–Nemhauser formulation. For $m = 2$, this procedure always produced the same local optimum, for any starting solution and for any value of k between 1 and 20. For $m = 3$, several local optima were obtained, but their criterion values are close together. The same is true for $m = 4$.

The question is now how to determine the best value of the number m of groups. It is obvious that, if more groups are added, the optimal criterion value will decrease. Cramer et al. (1970) used a simulation experiment to estimate the decreases that can be solely ascribed to enlarging the number of groups. They generated a number of random surveys among 45 "colorless" people and computed locally optimal solutions for these into two, three and four groups. If one goes from one to two groups, then an average of 42 percent of the total weight remains, with a very small variance, as compared to only 30 percent in case of the actual survey data. If more groups are added, the decreases for the simulation and for the survey are about the same. It was concluded that there is a clear distinction between two, but no more than two, groups.

5.4. Related models and applications

The relation of the *clique m-partitioning problem* to the graph coloring problem and the max cut problem has already been pointed out.

A variant of the model occurs if upper bounds on the group sizes are specified. Arnold et al. (1973) describe an application that arises when one is organizing a scientific meeting: nodes correspond to sessions, weights to conflicts between sessions, and colors to time slots. They used a combination of random sampling and local search.

Kernighan and Lin (1970) consider the problem of partitioning V into two equal-size subsets so as to minimize the total weight of the edges whose ends are in *different* subsets. They propose a variable-depth search method, which is a precursor of their algorithm for the traveling salesman problem (see section 1.3).

The *clique partitioning problem* serves as a model for a variety of clustering problems. Grötschel and Wakabayashi (1989) give several examples, such as the classification of animals with respect to morphological and behavioral characteristics, and the classification of member countries of the United Nations on the basis of voting behavior. All these problem situations can be captured under the general heading of the *aggregation of binary relations*, which is a central topic of interest in the area of qualitative data analysis.

6. Test cover

6.1. Recognizing diseases

In 1979, the CWI received a request from the Department of Mathematics at the Agricultural University in Wageningen: "Enclosed you will find a 0–1 matrix B with 63 rows and 28 columns. We define a 0–1 matrix A with 63 rows and 378 columns. Each column of A corresponds to a pair of columns of B (note that $378 = 28 \cdot 27/2$) and is obtained by adding those columns modulo 2. We would be interested in a solution to the set covering problem on A^T."

The *set covering problem* is here the problem of finding a minimum number of rows of A such that, in each column, at least one of these rows has a 1. The difficulty of solving large set covering problems as well as our professional curiosity motivated us to transform backwards. We identified the rows of B with tests, the columns with items, and each entry (i, j) with the result of test i applied to item j. The problem is then to find a minimum number of tests such that, for each pair of items, at least one of these tests distinguishes the two items. This is the *test cover problem*.

Before discussing these models and their relation in more detail, let us clarify the practical background. It turned out that tests and items were plant varieties and plant diseases, respectively. A minimum number of varieties that discriminates between all diseases should provide an efficient and economical setup for recognizing diseases. The problem arose as part of a project carried out at the Research Institute of Plant Protection in Wageningen in cooperation with the International Maize and Wheat Improvement Center (CIMMYT). We provided a program that incorporates the algorithm described in section 6.3. It computed an optimal solution of seven tests for the above instance (the best known solution used eight) and has been used successfully on many other instances.

6.2. *Model formulation*

Given is a finite set N and a family $\mathcal{T}$ of subsets of N; N contains the items and $\mathcal{T}$ is the collection of tests. A *test cover* is a subfamily $\mathcal{T}' \subset \mathcal{T}$ such that, for each pair $\{j, k\} \subset N$, there is a test $T \in \mathcal{T}'$ such that T distinguishes between j and k, i.e., $|T \cap \{j, k\}| = 1$. The problem is to find a test cover of minimum cardinality.

A more familiar problem is the set covering problem: given a finite set M and a family $\mathcal{S}$ of subsets of M, determine a minimum-size subfamily $\mathcal{S}' \subset \mathcal{S}$ for which $\bigcup_{S \in \mathcal{S}'} S = M$.

As suggested above, we can formulate each instance of the test cover problem as a set covering problem. We define an element in $\mathcal{M}$ for each pair of items in N, and we create a subset $S \in \mathcal{S}$ for each test $T \in \mathcal{T}$; the elements in S are precisely the pairs of items in N that are distinguished by T. It is immediate that a subfamily $\mathcal{S}' \subset \mathcal{S}$ covers M if and only if the corresponding subfamily $\mathcal{T}' \subset \mathcal{T}$ is a test cover.

Through a relation with an optimization problem on graphs, which we will briefly discuss in section 6.4, it turns out that the test cover problem is already NP-hard if $|T| \leqslant 2$ for all $T \in \mathcal{T}$.

6.3. *Solution approaches*

Any algorithm for set covering can be used for the test cover problem. However, the quadratic blowup is not encouraging, and it even appears that the resulting set covering instances are particularly hard ones. Although all NP-hard combinatorial optimization problems are polynomially equivalent, it generally pays off to develop an algorithm that is specific to the problem at hand.

The problem instance described in the introduction was solved to optimality by a straightforward combination of approximation and branch and bound. In the first phase, a greedy algorithm constructs a reasonable solution and a local search algorithm tries to improve on it. The greedy algorithm selects, at each step, the test that distinguishes the greatest number of pairs that are not yet distinguished by the tests selected so far. Iterative improvement then produces a good solution, the value of which serves as an initial upper bound for the branch and bound process. In this second phase, the tests $T \in \mathcal{T}$ are put in non-increasing order of "distinguishing power" $\min\{|T|, |N - T|\}$ and the subfamilies of $\mathcal{T}$ are enumerated in lexicographic order. Subfamilies are eliminated by a lower bound which is based on the simple observation that, for distinguishing n items, at least $\lceil \log_2 n \rceil$ tests are needed.

One of the first papers on the test cover problem is by Moret and Shapiro (1985). They investigate the worst-case behavior of approximation algorithms and the derivation of lower bounds. A negative result is that the greedy algorithm can perform as badly as for the general set covering problem and produce solutions that are off by a factor of $\Theta(\log|N|)$. The above project has inspired further work on approximation and optimization algorithms for the test cover problem.

6.4. Related models and applications

An interesting special case of the test cover problem occurs if all tests have cardinality 2. In terms of a graph G with node set N and edge set $\mathcal{T}$, a test cover is an edge subset $\mathcal{T}' \subset \mathcal{T}$ such that no two nodes in N have the same incidence relations with respect to $\mathcal{T}'$. This turns out to be equivalent to the requirement that the subgraph $G' = (N, \mathcal{T}')$ has no isolated edges and at most one isolated node. One can establish a strong relation between this problem and the problem of finding a maximum number of node-disjoint paths of length 2 in a graph. This implies NP-hardness of the restricted test cover problem and has interesting consequences for the worst-case behavior of approximation algorithms.

Moret and Shapiro (1985) mention applications of the test cover problem in fault testing and diagnosis, pattern recognition, and biological identification.

7. Bottleneck extrema

7.1. Locating obnoxious facilities

In a regional development plan, a number of sites has been identified at which residual quarters as well as industrial areas will be located. The industrial areas will accommodate obnoxious facilities. The problem is how to select sites for the industrial areas so as to minimize the inconvenience caused to the residual quarters.

Our presentation follows Hsu and Nemhauser (1979). We did not encounter the problem in practice, but we decided to include it here because it nicely illustrates the elegant theory of bottleneck extrema.

7.2. Model formulation

Let us assume that there are n sites, k of which will be industrial areas. We construct a graph $G = (V, E)$ with $V = \{1, \ldots, n\}$ and $\{i, j\} \in E$ if and only if sites i and j influence each other, i.e., if an industrial area at one of the sites would be a nuisance to a residual quarter at the other site. With each edge $e = \{i, j\} \in E$, a weight d_e is associated, representing the distance between sites i and j.

call that, for all $U \subset V$ with $U \neq \emptyset$ and $U \neq V$, $\delta(U)$ is the set of edges with e end in U. The set $\delta(U)$ is called a *cut*; its removal from G disconnects U and $V \backslash U$. If $|U| = k$, then $\delta(U)$ will be called a *k-cut*.

One way of minimizing annoyance is to maximize the minimum distance between any residential quarter and its closest industrial area. The problem is then to find a k-cut whose minimum edge weight is maximum:

$$\max_{U:|U|=k} \min_{e\in\delta(U)} d_e . \tag{7.1}$$

We present a polynomial-time algorithm for this problem in the next section.

7.3. *Solution approach*

The solution of (7.1) is based on the theory of bottleneck extrema developed by Edmonds and Fulkerson (1970). We outline some of their results below.

Let S be a finite set. A *clutter* $\mathscr{C}$ on S is a collection of subsets of S such that no set in $\mathscr{C}$ includes any other set in $\mathscr{C}$. The *blocker* $\mathscr{B}$ of $\mathscr{C}$ is the collection of subsets of $\mathscr{S}$ that intersect all sets in $\mathscr{C}$ and are minimal under inclusion. Note that $\mathscr{B}$ is again a clutter.

Edmonds and Fulkerson proved two duality results. First, $\mathscr{C}$ is the blocker of its blocker $\mathscr{B}$. Secondly, for any real-valued function f on S,

$$\max_{C\in\mathscr{C}} \min_{e\in C} f_e = \min_{B\in\mathscr{B}} \max_{e\in B} f_e . \tag{7.2}$$

They also presented a simple threshold algorithm for the max–min problem and an analogous dual threshold method for the min–max problem. The choice between these two algorithms depends on the relative efficiency of recognizing members of $\mathscr{C}$ and $\mathscr{B}$. We will describe the dual method.

Suppose that the elements of S are indexed so that $f_1 \leqslant f_2 \leqslant \cdots \leqslant f_{|S|}$. Suppose further that $e^* \in S$ is such that $\{1,\ldots,e^*-1\}$ does not include a member of $\mathscr{B}$ but that $\{1,\ldots,e^*\} \supset B^*$ for some $B^* \in \mathscr{B}$. Then B^* solves the min–max problem. By bisection search over S, e^* can be found in $O(\log|S|)$ iterations, where each iteration tests for inclusion of a member of $\mathscr{B}$.

We now return to problem (7.1). It suffices to consider only minimal k-cuts: if $E' \subset E''$, then $\min_{e\in E'} d_e \geqslant \min_{e\in E''} d_e$. We let S corresponded to the edge set E, f to the weight function d, and $\mathscr{C}$ to the collection of minimal k-cuts; note that $\mathscr{C}$ is a clutter. Problem (7.1) can now be stated as $\max_{C\in\mathscr{C}} \min_{e\in C} f_e$.

Rather than solving the recognition problem for members of $\mathscr{C}$, we will characterize its blocker $\mathscr{B}$ and give an $O(n^2)$ membership test for $\mathscr{B}$. We thus obtain an $O(n^2 \log n)$ algorithm for computing $\min_{B\in\mathscr{B}} \max_{e\in B} f_e$ and, by (7.2), for solving (7.1).

The crucial observation to make is that a subset $B \subset E$ intersects all k-cuts if and only if its complement $E\backslash B$ is not a k-cut. It immediately follows that $\mathscr{B}$ contains the minimal $B \subset E$ such that no collection of connected components of the graph $G_B = (V, B)$ contains exactly k nodes.

As to the membership test, consider some $E' \subset E$. Suppose $G_E{}' = (V, E')$ has m components and let component h have a_h nodes, for $h = 1,\ldots,m$. Then $E' \supset B$

for some $B \in \mathcal{B}$ if and only if

$$\sum_{h=1}^{m} a_h x_h = k ,$$
$$x_h \in \{0, 1\} \quad \text{for } h = 1, \ldots , m$$

has no solution. This can be tested by dynamic programming in $O(km) = O(n^2)$ time.

7.4. *Related models*

Discrete location theory is one of the cornerstones of combinatorial operations research. Its basic problem types are the following. In the *uncapacitated plant location problem*, one has to find a set of locations such that the sum of the setup costs for facilities and the transportation costs between customers and facilities is minimized. In the *k-median problem*, one has to locate k facilities so as to minimize the sum of the distances between each customer and its closest facility, while in the *k-center problem*, the maximum of these distances is to be minimized.

Mirchandani and Francis (1990) collected a number of survey articles on discrete location theory.

8. Minimum cost flow

8.1. *2-dimensional proportional representation*

Gooi en Vechtstreek is a region in the Netherlands, just east of Amsterdam. There is a regional council, the members of which are appointed by and from the participating local councils. The composition of the regional council should be a fair representation of the local interests as well as of the political views in the region.

Each local council has an odd number of members. It is prescribed that a quarter of them, rounded to the nearest integer, is appointed to the regional council. This should take care of a proportional representation of local interests. However, if the allocation of seats to political parties is left completely to the local councils, then there is an obvious danger that overall disproportionalities in the representation of political views will occur. Coordination is required.

It has been agreed that the chairman of the regional cooperation uses the outcome of the local elections to determine the number of seats to be allocated to each party from each local council. The result of his reflections has the status of an advice to the local councils. In current practice, he applies the method proposed by Anthonisse (1984) in arriving at his advice. This method is described below.

8.2. Model formulation

We first briefly consider the question of 1-dimensional proportional representation. This problem has attracted a lot of interest, in politics as well as in mathematics. It can be formulated as follows.

Suppose there are P parties, V votes, and S seats. Party p has received v_p votes, with $\sum_{p=1}^{P} v_p = V$. Ideally, party p should receive $s_p = Sv_p/V$ seats. However, the s_p are generally non-integral and have to be rounded. Let a bivariate function f be given, where $f(s_p, x_p)$ measures the distance between the ideal number of seats s_p and the actual allotment x_p. The problem is then to find $x_1, \ldots, x_P$ such that

$$\sum_{p=1}^{P} f(s_p, x_p)$$

is minimized subject to

$$\sum_{p=1}^{P} x_p = S, \tag{8.1}$$

$$x_p \in \mathbb{N} \cup \{0\} \quad \text{for } p = 1, \ldots, P.$$

Various methods for solving this problem have been proposed. Hamilton's method of the greatest remainders corresponds to $f(s_p, x_p) = |x_p - s_p|$, Jefferson's method of the greatest divisors to $f(s_p, x_p) = (x_p - (s_p - \frac{1}{2}))^2/s_p$, and Webster's method to $f(s_p, x_p) = (x_p - s_p)^2/s_p$. For a historical and mathematical overview of 1-dimensional proportional representation, we refer to Balinski and Young (1982). They list a number of properties a perfect method of apportionment should satisfy and show that no such method exists. They argue convincingly that Webster's method comes closest towards "meeting the ideal of one man, one vote".

As to the problem of 2-dimensional proportional representation, suppose there are P parties and M municipalities. In the local council of municipality m, party p has v_{mp} seats. The size of council m is $V_{m\cdot} = \sum_{p=1}^{P} v_{mp}$, the regional strength of party p is $V_{\cdot p} = \sum_{m=1}^{M} v_{mp}$, and $V = \sum_{m=1}^{M} \sum_{p=1}^{P} v_{mp}$. Let $w_m = \lfloor (V_{m\cdot} + 1)/4 \rfloor$ denote the number of members of council m that are to be appointed in the regional council, and let $S = \sum_{m=1}^{M} w_m$ be the total number of seats in the regional council. Ideally, party p should receive $s_p = SV_{\cdot p}/V$ regional seats, $t_{mp} = Sv_{mp}/V$ of which should come from municipality m. The decision variables are x_p, the actual allotment to party p, and y_{mp}, the number of members of party p to be appointed from the council of municipality m. In the formulations below, f can be any convex bivariate distance function.

The problem is solved in two stages. First, we find the x_p such that

$$\sum_{p=1}^{P} f(s_p, x_p)$$

is minimized subject to

$$\sum_{p=1}^{P} y_{mp} = w_m \quad \text{for } m = 1, \ldots, M\,, \tag{8.2}$$

$$\sum_{m=1}^{M} y_{mp} = x_p \quad \text{for } p = 1, \ldots, P\,, \tag{8.3}$$

$$y_{mp} \leqslant v_{mp} \quad \text{for } m = 1, \ldots, M\,,\ p = 1, \ldots, P\,, \tag{8.4}$$

$$\begin{aligned} &y_{mp} \in \mathbb{N} \cup \{0\} \quad \text{for } m = 1, \ldots, p = 1, \ldots, P\,, \\ &x_p \in \mathbb{N} \cup \{0\} \quad \text{for } p = 1, \ldots, P\,. \end{aligned} \tag{8.5}$$

Secondly, given the x_p, we find the y_{mp} such that

$$\sum_{m=1}^{M} \sum_{p=1}^{P} f(t_{mp}, y_{mp})$$

is minimized subject to (8.2)–(8.5).

The first problem obviously has a feasible solution: any sample of w_m from the $V_{m\cdot}$ members satisfies (8.2), (8.4) and (8.5), and (8.3) then defines the x_p. Given these x_p, the second problem is also feasible, as the constraints remain the same. The reader may wonder why we have not simplified the first stage by relaxing (8.2)–(8.5) into (8.1). The reason is that this may yield an infeasible second stage.

Both problems can be modeled in terms of minimum cost flows (see chapter 2) and as such be solved in polynomial time. At the first stage, we define a network with a source α, a node m for each municipality, a node p for each party, and a sink ω. There are arcs (α, m) with given flow values w_m, arcs (m, p) with capacities v_{mp}, and unconstrained arcs (p, ω); $f(s_p, x_p)$ denotes the cost of sending x_p units of flow through the arc (p, ω). A feasible solution $((x_p), (y_{mp}))$ now corresponds to a flow of value S from α to ω, with flow values y_{mp} through the arcs (m, p) and x_p through (p, ω). The conditions (8.2) and (8.3) are flow conservation constraints at the nodes m and p; the conditions (8.4) represent the capacity constraints of the arcs (m, p). We have to find a feasible flow that minimizes the total costs of the flows x_p through the arcs (p, ω). At the second stage, we use the same network, except that the arcs (m, p) have costs $f(t_{mp}, y_{mp})$ and the arcs (p, ω) have given flow values x_p. We now have to find feasible flows y_{mp} through the arcs (m, p) of minimum total costs.

8.3. Solution approaches

Minimum cost flow problems are treated in depth in chapter 2. Minoux (1986) gives a polynomial-time algorithm for finding minimum cost integral flows with separable convex cost functions.

8.4. Related models

Following Anthonisse's work, Balinski and Demange (1989) pursued an axiomatic approach to two problems, one of which generalizes the problem considered above. Given are a nonnegative matrix $V=(v_{mp})$ and a positive integer S; one may give v_{mp} and S the same interpretation as before. In addition, nonnegative integers $w_m^-, w_m^+, x_p^-, x_p^+$ are given. An *allocation* is defined as a matrix $Y=(y_{mp})$ of the same dimensions as V with $w_m^- \leq \sum_p y_{mp} \leq w_m^+$, $x_p^- \leq \sum_m y_{mp} \leq x_p^+$, and $\sum_m \sum_p y_{mp} = S$. What should it mean to say that an allocation Y is proportional to V? And what should this mean for an *apportionment*, i.e., an *integer* allocation?

9. Interval scheduling

9.1. Dealing with nasty clients

A Dutch firm, primarily engaged in the retail trade, had decided to diversify and had acquired a large number of summer cottages. A client can make a reservation at any one of the firm's branches and is immediately told whether a cottage is still available for the period (s)he is applying for. Only at a later stage is it determined in which cottage each accepted client will spend the holidays. This procedure gave rise to a number of questions.

Does there exist a simple rule that indicates whether a client can be accepted? Yes, there does, as we will clarify below: cottages can be assigned to clients in their desired periods if and only if, at any time, the number of clients is no greater than the number of cottages. How about a method that assigns the accepted clients to a minimum number of cottages? This exists as well: assign the clients to cottages in order of their starting times, giving priority to cottages used before.

Both questions could be answered during the first contact with the firm's employee who sought the advice of the CWI. All seemed well until, while leaving, a trivial complication crossed his mind: a client can reserve a specific cottage by paying Hfl. 25 upon application and is then preassigned. This appears to have a dramatic effect on the problem's computational complexity. The above necessary and sufficient condition for acceptance remains valid only under the assumption that the clients would be willing to move into another cottage now and then. But under the more realistic assumption that these people do not want to move when they are on holiday, the problem turns out to be NP-complete.

We never heard from our client again. The complications caused by the nasty clients are probably trivial indeed and do not prohibit the application of the existing methods.

The above account follows Anthonisse and Lenstra (1984). We will deal with the technical details below.

9.2. Model formulation

We rephrase the problem in scheduling terminology. Cottages will be represented by machines and clients by jobs. There are m identical parallel machines, each of which can handle at most one job at a time. There are n independent jobs j, which need processing on one of the machines during the time interval (s_j, t_j) $(j = 1, \dots, n)$. First of all, we are interested in the minimum number of machines needed to process all jobs. The solution of this problem goes back to Dantzig and Fulkerson (1954).

Let us define a *partial order* $\rightarrow$ on the set of jobs. We say that $j \rightarrow k$ whenever $t_j \leq s_k$, i.e., when job j is completed before job k starts. A *chain* in the job set is a subset $\{j_1, j_2, \dots, j_k\}$ with $j_1 \rightarrow j_2 \rightarrow \cdots \rightarrow j_k$. The jobs in a chain can be consecutively scheduled on one machine, and, conversely, any schedule on one machine corresponds to a chain in the job set. The minimum number of machines needed to process all jobs is therefore equal to the minimum number of chains into which the job set can be partitioned.

Jobs j and k are *unrelated* if neither $j \rightarrow k$ nor $k \rightarrow j$. An *antichain* is a set of pairwise unrelated jobs. Any two jobs in antichain overlap in time; by a property of intervals, this is equivalent to saying that all jobs in an antichain overlap at a certain time.

We now invoke Dilworth's chain decomposition theorem: for every partially ordered set, the minimum number of chains needed to cover all elements is equal to the maximum number of elements in an antichain. Or, the minimum number of machines needed to process all jobs is equal to the maximum number of jobs that require simultaneous processing. For fast algorithms that actually assign the jobs to a minimum number of machines, we refer to section 9.3.

Another way of modeling the interval scheduling problem is in terms of *interval graphs*. Associated with the job set, we define the interval graph $G = (\{1, \dots, n\}, E)$ where $\{j, k\} \in E$ if and only if jobs j and k overlap in time. We recall that a *clique* in a graph is a subset of pairwise adjacent nodes, a *stable set* is a subset of pairwise non-adjacent nodes, and the *chromatic number* is the smallest k for which the node set can be partitioned into k stable sets. Clearly, a clique of G corresponds to an antichain in the partially ordered set, a stable set of G corresponds to a chain, and the chromatic number of G is equal to the minimum number of machines that can accomodate all jobs.

For interval graphs, it is true in general that the chromatic number is equal to the maximum clique size. This result parallels Dilworth's decomposition theorem for partially ordered sets. A minimum coloring and a maximum clique in an interval graph can be found in polynomial time. We refer to chapter 4 for details.

We now turn to the situation in which some clients have been preassigned to specific cottages during certain periods. Machines will now correspond to maximal idle periods of the cottages and jobs to assigned clients. Machine i is available during the interval (a_i, b_i) $(i = 1, \dots, m)$; job j requires processing during the interval (s_j, t_j) $(j = 1, \dots, n)$. Note that, in contrast to the previous problem, the machines are not identical anymore. The question whether a client can be accepted boils down to the following problem: is it possible to pack the intervals (s_j, t_j) into the intervals (a_i, b_i)?

We may try to generalize the partial order model to this situation. We introduce dummy jobs $n + i$ requiring processing in $(-\infty, a_i)$ and $n + m + i$ requiring processing in (b_i, ∞), for $i = 1, \ldots, m$. Again, we write $j \rightarrow k$ if and only if job j is completed before job k starts. A feasible schedule corresponds to a decomposition of the job set into m chains, where the ith chain starts with job $n + i$ and ends with job $n + m + i$ $(i = 1, \ldots, m)$. Conversely, because $\{n + 1, \ldots, n + m\}$ and $\{n + m + 1, \ldots, n + 2m\}$ are antichains, any chain in a decomposition of the job set into m chains must start at some $n + i$ $(1 \leq i \leq m)$ and end at some $n + m + i'$ $(1 \leq i' \leq m)$. Unfortunately, there is nothing to guarantee that $i = i'$, and therefore a chain does not necessarily correspond to a schedule on one machine. It is not hard to see, however, that from such a chain decomposition a *preemptive* schedule can be constructed, in which the processing of a job may be interrupted on one machine and continued on another machine.

The nonpreemptive problem appears to be much harder. Kolen et al. (1991) give a polynomial-time algorithm for the case of fixed m. They also prove that the general case is NP-complete, by relating the problem to a generalization of interval graph coloring.

9.3. *Solution approaches*

We restrict ourselves here to algorithms for the interval scheduling problem on identical machines.

Ford and Fulkerson (1962) give a simple $O(n^2)$ algorithm for decomposing a partially ordered set into chains. This so-called *staircase rule* finds a minimum number of chains in the case that the elements can be numbered so that $j \leq k$ implies that all predecessors of j are included in those of k. This condition holds for the partial order defined on the job set. The rule works as follows. Find the smallest element, in terms of the numbering. Repeatedly, find the smallest successor of the last element found, until no successor exists. Delete the chain that has been found, and repeat the process.

Gupta et al. (1979) give an $O(n \log n)$ algorithm, which builds the chains in parallel rather than in series. This rule, which was informally stated in section 9.1, is as follows. Put all of the machines on a stack S of idle machines. Order the s_j and t_j $(j = 1, \ldots, n)$ in nondecreasing order, where a t_j precedes an s_k in case of a tie; this yields a nondecreasing sequence $u_1, u_2, \ldots, u_{2n}$. Then, for $k = 1, \ldots, 2n$, do the following: if u_k corresponds to s_j, then assign job j to the machine on top of S, and remove this machine from S; if u_k corresponds to t_j, then put the machine to which job j was assigned on top of S.

9.4. *Related models*

Several generalizations of the interval scheduling problem on identical parallel machines have been investigated.

Arkin and Silverberg (1987) analyze the case in which there is a weight associated with each job and a maximum-weight subset of jobs that can be scheduled on m machines is to be found. They develop an $O(n^2 \log n)$ algorithm.

Fischetti et al. (1987, 1989) consider two problems types. In the first one, each machine is available for a period of length b, which starts at the starting time of the first job assigned to it. In the second problem type, each machine can perform no more than b time units of processing. In both cases, the number of machines is to be minimized. They show that both problems are NP-hard and developed branch and bound algorithms for their solution.

Another extension involves hierarchies of machines and jobs. There are m classes of machines and m classes of jobs. All machines are available during the same time interval. A job in class i requires processing during a given interval and can only be assigned to machines in classes $1, \ldots, i$. Does there exist a feasible schedule? Kolen et al. (1991) give a polynomial-time algorithm for the case that $m = 2$, using network flow techniques, and show that the case $m = 3$ is NP-complete. Subsequent work concerns optimization versions of this problem, where costs are associated with the machines.

10. Job shop scheduling

10.1. Production planning

Combinatorial optimization problems that arise in production planning tend to be both difficult to formulate and difficult to solve. That is, the problem is often characterized by constraints that are very specific to the situation at hand, and it is usually an easy matter to find many independent reasons for its NP-hardness. These observations may explain why the development and application of general software in the area of production planning is not nearly at the stage at which it is in vehicle routing.

In order to avoid complicating details, we have chosen to consider a standard problem type, the job shop scheduling problem, which is at the core of many practical production planning situations. It is described as follows. Given are a set of jobs and a set of machines. Each machine can handle at most one job at a time. Each job consists of a chain of operations, each of which needs to be processed during an uninterrupted time period of a given length on a given machine. The purpose is to find a schedule, i.e., an allocation of the operations to time intervals on the machines, that has minimum length.

This problem allows a number of relatively straightforward mathematical formulations. In addition, it is extremely difficult to solve to optimality. This is witnessed by the fact that a problem instance with only ten jobs, ten machines and one hundred operations, published in 1963, remained unresolved until 1986.

10.2. Model formulation

Given are a set $\mathcal{J}$ of jobs, a set $\mathcal{M}$ of machines, and a set $\mathcal{O}$ of operations. For each operation $i \in \mathcal{O}$, there is a job $J_i \in \mathcal{J}$ to which it belongs, a machine $M_i \in \mathcal{M}$ on which it requires processing, and a processing time $p_i \in \mathbb{N}$. There is a binary relation $\rightarrow$ on $\mathcal{O}$ that decomposes $\mathcal{O}$ into chains corresponding to the jobs; more specifically, if $i \rightarrow j$, then $J_i = J_j$ and there is no $k \not\in \{i, j\}$ with $i \rightarrow k$ or $k \rightarrow j$. The problem is to find a starting time S_i for each operation $i \in \mathcal{O}$ such that

$$\max_{i \in \mathcal{O}} S_i + p_i \tag{10.1}$$

is minimized subject to

$$S_i \geq 0 \qquad \text{for } i \in \mathcal{O}\,, \tag{10.2}$$

$$S_j - S_i \geq p_i \qquad \text{whenever } i \rightarrow j\,,\ i, j \in \mathcal{O}\,, \tag{10.3}$$

$$S_j - S_i \geq p_i \vee S_i - S_j \geq p_j \qquad \text{whenever } M_i = M_j\,,\ i, j \in \mathcal{O}\,. \tag{10.4}$$

The objective function (10.1) represents the schedule length, in view of (10.2). The conditions (10.3) are the job precedence constraints. The conditions (10.4) represents the machine capacity constraints, which make the problem NP-hard.

To obtain an *integer programming* formulation, we choose an upper bound T on the optimum and introduce a 0–1 variable y_{ij} for each ordered pair (i, j) with $M_i = M_j$, where $y_{ij} = 0$ ($y_{ij} = 1$) corresponds to $S_j - S_i \geq p_i$ ($S_i - S_j \geq p_j$). We now replace (10.4) by

$$\left.\begin{array}{l} y_{ij} \in \{0, 1\}\,, \\ y_{ij} + y_{ji} = 1\,, \\ S_i + p_i - S_j - Ty_{ij} \leq 0\,, \end{array}\right\} \text{ whenever } M_i = M_j\,,\ i, j \in \mathcal{O}\,. \tag{10.4$'$}$$

This formulation is closely related to the *disjunctive graph*. The disjunctive graph $G = (\mathcal{O}, A, E)$ has a node set $\mathcal{O}$, an arc set $A = \{(i, j) \mid i \rightarrow j\}$, and an edge set $E = \{\{i, j\} \mid M_i = M_j\}$; note that the arcs are directed and the edges are undirected. A weight p_i is associated with each node i. There is an obvious one-to-one correspondence between feasible values of the y_{ij} in (10.1), (10.2), (10.3), (10.4′) and orientations of the edges in E for which the resulting digraph is acyclic. Given any such orientation, we can determine feasible starting times by setting each S_i equal to the weight of a maximum-weight path in the digraph finishing at i minus p_i; the objective value is equal to the maximum path weight in the digraph. The problem is now to find an orientation of the edges in E that minimizes the maximum path weight.

10.3. Solution approaches

Optimization algorithms for job shop scheduling proceed by branch and bound. A node in the search tree is usually characterized by an orientation of each edge in a certain subset $E' \subset E$. The question is then how to compute a lower bound on the value of all completions of this partial solution.

A trivial lower bound is obtained by simply disregarding $E \backslash E'$ and computing the maximum path weight in the digraph $(\mathcal{O}, A \cup E')$. A more sophisticated bound is based on the relaxation of the capacity constraints of all machines except one: a machine $M' \in \mathcal{M}$ is selected, and the job shop problem is solved on the disjunctive graph $(\mathcal{O}, A \cup E', \{\{i, j\} \mid M_i = M_j = M'\})$. This reduces to a *single-*

machine problem, where the arcs in $A \cup E'$ define release and delivery times for the operations that are to be scheduled on M'. Although it is an NP-hard problem, there exist fairly efficient algorithms for its solution. The single-machine bound generalizes all previously proposed bounds (Lageweg et al. 1977). More recent (and more complicated) bounds use *surrogate duality relaxation* and *polyhedral techniques*.

A variety of branching schemes to generate the search tree and elimination rules to truncate it is available. For this and for more information on the lower bounds, we refer to Lawler et al. (1993).

Most *approximation algorithms* for job shop scheduling use a dispatch rule, which schedules the operations according to some priority function. Adams et al. (1988) developed a *sliding bottleneck heuristic*, which employs an ingenious combination of schedule construction and iterative improvement, guided by solutions to single-machine problems of the type described above. They also embedded this method in a second heuristic that proceeds by partial enumeration of the solution space.

Van Laarhoven et al. (1992) applied the principle of *simulated annealing* to the job shop scheduling problem. This is a randomized variant of iterative improvement. It is based on local search, but accepts deteriorations with a small and decreasing probability in the hope of avoiding bad optima and getting settled in a global optimum. In the present case, the neighborhood of a schedule contains all schedules that can be obtained by interchanging two operations i and j for which $M_i = M_j$ and the arc (i, j) is on a longest path.

The *computational merits* of all these algorithms are accurately reflected by their performance on the notorious 10-job 10-machine problem instance dating back to 1963.

The single-machine bound, maximized over all machines, has a value of 808. In 1975, McMahon and Florian used the single-machine bound and a branching scheme that constructs all left-justified schedules to arrive at a schedule of length 972, without proving optimality. In 1983, Fisher, Lageweg, Lenstra and Rinnooy Kan applied surrogate duality relaxation to find a lower bound of 813; the time requirements involved did not encourage them to carry on the search beyond the root of the tree. In 1984, Lageweg developed an improved implementation of the McMahon–Florian algorithm, with an adaptive search strategy, and found a schedule of length 930; he also computed a number of multi-machine bounds, ranging from a three-machine bound of 874 to a six-machine bound of 907. Two years later, Carlier and Pinson (1989) proved optimality of the value 930; they used a relaxation of the single-machine bound, a drastically different branching scheme, and many elimination rules. The main drawback of all these enumerative methods, aside from the limited problem sizes that can be handled, is their sensitivity to particular problem instances and even to the initial value of the upper bound.

The computational experience with polyhedral techniques that has been reported in recent years is slightly disappointing in view of what has been achieved for the traveling salesman problem and the linear ordering problem. However, the investigations in this direction are still at an initial stage.

Dispatch rules show an erratic behavior. The rule proposed by Lageweg et al. (1977) constructs a schedule of length 1082, and most other priority functions do worse. Adams et al. (1988) report that their sliding bottleneck heuristic obtains a schedule of length 1015 in ten CPU seconds, solving 249 single-machine problems on the way. Their partial enumeration procedure succeeds in finding the optimum, after 851 seconds and 270 runs of the first heuristic.

Five runs of the simulated annealing algorithm with a standard setting of the cooling parameters take 6000 seconds on average and procedure an average schedule length of 942.4, with a minimum of 937. If 6000 seconds are spent on deterministic neighborhood search, which accepts only true improvements, then more than 9000 local optima are found, the best one of which has a value of 1006. Five runs with a much slower cooling schedule take about 16 hours each and produce solution values of 930 (twice), 934, 935 and 938. In comparison to other approaches, simulated annealing requires unusual computation times, but it yields consistently good solutions with a modest amount of human implementation effort and relatively little insight into the combinatorial structure of the problem type under consideration.

10.4. Related models

The theory of scheduling is concerned with the optimal allocation of scarce resources to activities over time. It has been the subject of extensive research over the past decades. The emphasis has been on the investigation of *deterministic machine scheduling* problems, in which each activity requires at most one resource at a time, each resource can perform at most one activity at a time, and all problem data are known in advance. This problem class is surveyed extensively by Lawler et al. (1993). The results for these problems have reached the level of detail that a computer program is being used to maintain a record of the complexity status of thousands of problem types.

We mention two natural extensions of this class that are of obvious practical importance. In *resource-constrained project scheduling*, an activity may require several resources to be performed and a resource may be able to handle several activities simultaneously. In *stochastic scheduling*, some problem parameters are random variables. Either class has generated an impressive literature of its own; see Lawler et al. (1993) for references.

Acknowledgements

We gratefully acknowledge the valuable comments and suggestions of David Shmoys and the editors. The research of the second author was partially supported by NSF Grant CCR-89-96272 with matching funds from UPS and Sun, and by the Cornell Computational Optimization Project.

References

Adams, J., E. Balas and D. Zawack

[1988] The shifting bottleneck procedure for job shop scheduling, *Management Sci.* **34**, 391–401.

Anthonisse, J.M.

[1984] Proportional representation in a regional council, *CWI Newsletter* 5, 22–29.

Anthonisse, J.M., and J.K. Lenstra

[1984] Operational operations research at the Mathematical Centre, *European J. Oper. Res.* **15**, 293–296.

Anthonisse, J.M., J.K. Lenstra and M.W.P. Savelsbergh

[1988] Behind the screen: DSS from an OR point of view, *Decision Support Systems* **4**, 413–419.

Arkin, E.M., and E.B. Silverberg

[1987] Scheduling jobs with fixed start and end times, *Discrete Appl. Math.* **18**, 1–8.

Arnold, L.R., R.E. Beckwith and C.M. Jones

[1973] Scheduling the 41st-ORSA-Meeting sessions: the visiting-fireman problem, II, *Oper. Res.* **21**, 1095–1103.

Balinski, M.L., and G. Demange

[1989] An axiomatic approach to proportionality between matrices, *Math. Oper. Res.* **14**, 700–719.

Balinski, M.L., and H.P. Young

[1982] *Fair Representation: Meeting the Ideal of One Man, One Vote* (Yale University Press, New Haven, CT).

Bland, R.G., and D.F. Shallcross

[1989] Large traveling salesman problems arising from experiments in X-ray crystallography: a preliminary report on computation, *Oper. Res. Lett.* **8**, 125–128.

Bodin, L., B. Golden, A. Assad and M. Ball

[1983] Routing and scheduling of vehicles and crews: the state of the art, *Comput. Oper. Res.* **10**, 63–211.

Carlier, J., and E. Pinson

[1989] An algorithm for solving the job-shop problem, *Management Sci.* **35**, 164–176.

Carlson, R.C., and G.L. Nemhauser

[1966] Scheduling to minimize interaction cost, *Oper. Res.* **14**, 52–58.

Christofides, N.

[1985] Vehicle routing, in: *The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization*, eds. E.L. Lawler, J.K. Lenstra, A.H.G. Rinnooy Kan and D.B. Shmoys (Wiley, Chichester) ch. 12, pp. 431–448.

Cramer, J.S., A.I.M. Kool, J.K. Lenstra and G. De Leve

[1970] *Beroep en Opleiding – Een Enquête onder Econometristen* (in Dutch: *Profession and Education: a Survey Among Econometricians*) (Institute for Actuarial Sciences and Econometrics, University of Amsterdam).

Dantzig, G.B., and D.R. Fulkerson

[1954] Minimizing the number of tankers to meet a fixed schedule, *Naval Res. Logist. Quart.* **1**, 217–222.

Desrochers, M., J.K. Lenstra, M.W.P. Savelsbergh and F. Soumis

[1988] Vehicle routing with time windows: optimization and approximation, in: *Vehicle Routing: Methods and Studies*, eds. B.L. Golden and A.A. Assad (North-Holland, Amsterdam) pp. 65–84.

Edmonds, J.R., and D.R. Fulkerson

[1970] Bottleneck extrema, *J. Combin. Theory* **8**, 299–306.

Fischetti, M., S. Martello and P. Toth

[1987] The fixed job schedule problem with spread-time constraints, *Oper. Res.* **35**, 849–858.

[1989] The fixed job schedule problem with working time constraints, *Oper. Res.* **37**, 395–403.

Fisher, M.L., and R. Jaikumar

[1981] A generalized assignment heuristic for vehicle routing, *Networks* **11**, 109–124.

Fisher, M.L., R. Jaikumar and L. Van Wassenhove

[1986] A multiplier adjustment method for the generalized assignment problem, *Management Sci.* **32**, 1095–1103.

Ford Jr, L.R., and D.R. Fulkerson

[1962] *Flows in Networks* (Princeton University Press, Princeton, NJ).

Grötschel, M., and O. Holland

[1991] Solution of large-scale symmetric travelling salesman problems, *Math. Programming* **51**, 141–202.

Grötschel, M., and Y. Wakabayashi

[1989] A cutting plane algorithm for a clustering problem, *Math. Programming* **45**, 59–96.

Guan, Meigu (Kwan Mei-Ko)

[1962] Graphic programming using odd or even points, *Chinese Math.* **1**, 273–277.

Gupta, U.I., D.T. Lee and J.Y.-T. Leung

[1979] An optimal solution for the channel-assignment problem, *IEEE Trans. Comput.* **C-28**, 807–810.

Held, M., and R.M. Karp

[1970] The traveling-salesman problem and minimum spanning trees, *Oper Res.* **18**, 1138- 1162.

[1971] The traveling-salesman problem and minimum spanning trees: Part II, *Math. Programming* **1**, 6–25.

Hsu, W.-L., and G.L. Nemhauser

[1979] Easy and hard hottleneck location problems, *Discrete Appl. Math.* **1**, 209–215.

Johnson, D.S., and C.H. Papadimitriou

[1985] Performance guarantees for heuristics, in: *The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization*, eds. E.L. Lawler, J.K. Lenstra, A.H.G. Rinnooy Kan and D.B. Shmoys (Wiley, Chichester) ch. 5, pp. 145–180.

Jünger, M.

[1985] *Polyhedral Combinatorics and the Acyclic Subdigraph Problem* (Heldermann, Berlin).

Kernighan, B.W., and S. Lin

[1970] An efficient heuristic procedure for partitioning graphs, *Bell System Tech. J.* **49**, 291–307.

Knuth, D.E., and M.F. Plass

[1981] Breaking paragraphs into lines, *Software – Practice and Experience* **11**, 1119–1184.

Kolen, A.W.J., J.K. Lenstra and C.H. Papadimitriou

[1991] *Interval scheduling problems*, unpublished Manuscript.

Lageweg, B.J., J.K. Lenstra and A.H.G. Rinnooy Kan

[1977] Job-shop scheduling by implicit enumeration, *Management Sci.* **24**, 441–450.

Lawler, E.L.

[1976] *Combinatorial Optimization: Networks and Matroids* (Holt, Rinehart & Winston, New York).

Lawler, E.L., J.K. Lenstra, A.H.G. Rinnooy Kan and D.B. Shmoys

[1985] eds., *The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization* (Wiley, Chichester).

[1993] Sequencing and scheduling: algorithms and complexity, in: *Logistics of Production and Inventory, Handbooks in Operations Research and Management Science*, Vol. 4, eds. S.C. Graves, A.H.G. Rinnooy Kan and P. Zipkin (North-Holland, Amsterdam) pp. 445–522.

Lenstra, J.K.

[1974] Clustering a data array and the traveling-salesman problem, *Oper. Res.* **22**, 413–414.

Lenstra, J.K., and A.H.G. Rinnooy Kan

[1975] Some simple applications of the travelling salesman problem, *Oper. Res. Quart.* **26**, 717–733.

[1981] Complexity of vehicle routing and scheduling problems, *Networks* **11**, 221–227.

Lin, S.

[1965] Computer solutions of the traveling salesman problem, *Bell System Tech J.* **44**, 2245–2269.

Lin, S., and B.W. Kernighan

[1973] An effective heuristic algorithm for the traveling salesman problem, *Oper. Res.* **21**, 498–516.

McCormick Jr, W.T., P.J. Schweitzer and T.W. White

[1972] Problem decomposition and data reorganization by a clustering technique, *Oper. Res.* **20**, 993–1009.

Minoux, M.

[1986] Solving integer minimum cost flows with separable convex cost objective polynomially, *Math. Programming Stud.* **26**, 237–239.

Mirchandani, P.B., and R.L. Francis
[1990] eds., *Discrete Location Theory* (Wiley, New York).

Moret, B.M.E., and H.D. Shapiro
[1985] On minimizing a set of tests, *SIAM J. Sci. Statist. Comput.* **6**, 983–1003.

Padberg, M.W., and G. Rinaldi
[1987] Optimization of a 532-city symmetric traveling salesman problem by branch and cut, *Oper Res. Lett.* **6**, 1–7.

Ratliff, H.D., and A.S. Rosenthal
[1983] Order picking in a rectangular warehouse: A solvable case of the traveling salesmán problem, *Oper. Res.* **31**, 507–521.

Reinelt, G.
[1985] *The Linear Ordering Problem. Algorithms and Applications* (Heldermann, Berlin).

Savelsbergh, M.W.P.
[1992] *Computer Aided Routing*, CWI Tract No. 75 (Centre for Mathematics and Computer Science, Amsterdam).

Van Laarhoven, P.J.M., E.H.L. Aarts and J.K. Lenstra
[1992] Job shop scheduling by simulated annealing, *Oper. Res.* **40**, 113–125.

CHAPTER 36

Combinatorics in Electrical Engineering and Statics

András RECSKI

Technical University of Budapest, Faculty of Electrical Engineering and Informatics, Department of Mathematics and Computer Science, Budapest 1521, Hungary

Contents

HANDBOOK OF COMBINATORICS
Edited by R. Graham, M. Grötschel and L. Lovász

In this chapter we present three engineering problems where combinatorial methods are needed for the solution. In order to emphasize the *methods* we (i) proceed in the solutions of the three problems simultaneously, and (ii) do not intend to present the engineering problems in their most general form.

Previous surveys of similar character are Iri (1983), Iri and Fujishige (1981), Murota (1987), Recski (1984), Roth (1981) and Sugihara (1986). These and other ideas in a more detailed form can be found in a monograph of Recski (1989). See also the forthcoming monograph of Crapo and Whiteley (1994).

The presented applications refer to electric network theory, to statics and to pattern recognition. Further applications of similar combinatorial tools are also known in control and system theory, see Iri and Fujishige (1981) and in geodesy (Spriggs and Snay 1982).

1. The problems

An *electric network* is the interconnection of the following types of devices:

(i) *voltage sources* whose voltage u is a given function of time and whose current i is arbitrary,

(ii) *current sources* with i given and u arbitrary,

(iii) *linear n-ports* (or *multiports*) with a given number n of node-pairs (called ports) so that the voltages $u_1, u_2, \ldots, u_n$ and the currents $i_1, i_2, \ldots, i_n$ of the ports are related by

$$A\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} + B\begin{pmatrix} i_1 \\ i_2 \\ \vdots \\ i_n \end{pmatrix} = \mathbf{0}\,, \qquad (*)$$

where $\boldsymbol{A}, \boldsymbol{B}$ are real matrices of size $n \times n$ with $r(\boldsymbol{A}\,|\,\boldsymbol{B}) = n$.

For example, an Ohmic resistor is a 1-port with $u - Ri = 0$, or an ideal transformer is a 2-port with $u_1 - ku_2 = 0$, $ki_1 + i_2 = 0$.

The interconnection of devices is described by a graph whose edges correspond to voltage or current sources, or to ports of the multiports. Figure 1.1 shows a fictitious network and the corresponding network graph. The symbols 1, 2 and 3 correspond to a voltage source, to a current source and to a resistor, respectively while (4), (5) and (6), (7), (8) are the ports of a 2-port and a 3-port, respectively.

The signed sum of voltages along a circuit of the network graph is zero (for example, $u_5 = u_7$ on fig. 1.1) by Kirchhoff's voltage law. Similarly the signed sum of currents along a cut set is zero (Kirchhoff's current law). A network is *uniquely solvable* if all the voltages and currents can uniquely be expressed as the functions of the voltages of the voltage sources and the currents of the current sources, using Kirchhoff's voltage and current laws and the multiport equations of form (*).

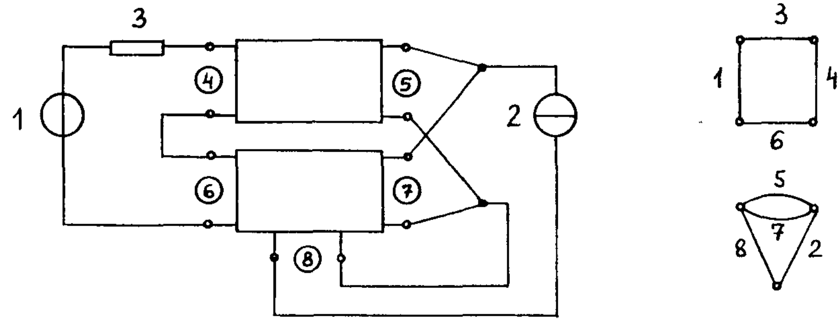


Figure 1.1.

Problem 1.1. Decide whether an electric network N, composed of voltage and current sources and multiports, has a unique solution.

A *framework* is a set of rigid bars interconnected by rotatable joints. Intuitively, call a framework rigid if every motion of it is a congruence. (The exact definition of rigidity is postponed until the next section.) For example, the first system of fig. 1.2 is rigid, the second one is not, while the third one is rigid in the 2-dimensional space and nonrigid in the 3-dimensional one.

The nonrigid examples in fig. 1.2 are "mechanisms": Fixing appropriate joints (to avoid congruent motions of the whole system) some other joints can have a continuous motion. However, the framework of fig. 1.3 will also be considered as nonrigid since the solid joint has an "infinitesimal" motion even if all the other joints are fixed.

Problem 1.2. Decide whether a framework F is rigid.

How can one reconstruct a convex polyhedron if only its projection (say, from above) is known as a 2-dimensional drawing with straight line segments? The

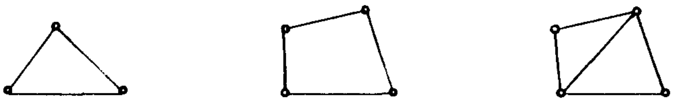


Figure 1.2.

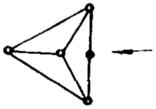


Figure 1.3.

drawing is defined as the projected picture itself *plus* the sets of vertices, edges and faces, with given lists of their incidences.

Problem 1.3. Decide whether a drawing D in the xy-plane arises as the 2-dimensional projection of a 3-dimensional convex polyhedron.

2. Are these problems linear?

In case of Problem 1.1 the answer is clearly in the affirmative. Kirchhoff's equations and the multiport equations of form (*) can be collected into a single system $\boldsymbol{W}\boldsymbol{t} = \boldsymbol{0}$ where $\boldsymbol{t}^{\mathrm{T}} = (u_1, u_2, \ldots, i_1, i_2, \ldots)$. Hence the network N is uniquely solvable if and only if the above coefficient matrix $\boldsymbol{W}$ is nonsingular.

In case of Problem 1.2 the linearity is not obvious. Let us restrict ourselves to the 2-dimensional case and let (x_i, y_i) be the coordinates of joint i. A rod between joints i and j means that

$$\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} = c_{ij} \text{ (a constant)}, \tag{2.1}$$

hence in case of m rods among n joints we obtain m *quadratic* equations among $2n$ unknowns. However, the time derivative of the square of eq. (2.1) is

$$(x_i - x_j)(\dot{x}_i - \dot{x}_j) + (y_i - y_j)(\dot{y}_i - \dot{y}_j) = 0 \tag{2.2}$$

and thus the collection of these equations can be written in the form $\boldsymbol{W}\boldsymbol{t} = \boldsymbol{0}$, where $\boldsymbol{t}^{\mathrm{T}} = (\dot{x}_1, \dot{x}_2, \ldots, \dot{y}_1, \dot{y}_2, \ldots)$ and the entries of $\boldsymbol{W}$ depend on the coordinates $x_1, x_2, \ldots, y_1, y_2, \ldots$ only, not on their time-derivatives.

Congruent motions of the whole plane (translations, rotations) are nontrivial solutions of this system and form a subspace of dimension 3. Hence the $m \times 2n$ matrix $\boldsymbol{W}$ cannot have more than $2n - 3$ linearly independent columns. The framework is defined to be *rigid* if $r(\boldsymbol{W}) = 2n - 3$.

In case of 3-dimensional frameworks, eq. (2.2) is replaced by

$$(x_i - x_j)(\dot{x}_i - \dot{x}_j) + (y_i - y_j)(\dot{y}_i - \dot{y}_j) + (z_i - z_j)(\dot{z}_i - \dot{z}_j) = 0, \tag{2.3}$$

and rigidity means that the matrix $\boldsymbol{W}$, now with size $m \times 3n$, has rank $3n - 6$.

The reader may verify that the framework of fig. 1.3 as well as the "mechanisms" of fig. 1.2 are nonrigid by this definition, justifying our previous remark on the "infinitesimal" motions. For the same reason, some authors call this concept infinitesimal rigidity; for brevity we use rigidity.

In case of Problem 1.3 recall that a vertex (x_i, y_i, z_i) of a polyhedron P is incident to a face F_j of P if and only if

$$a_j x_i + b_j y_i + z_i + d_j = 0, \tag{2.4}$$

where the plane of F_j is given by $\alpha_j x + \beta_j y + \gamma_j z + \delta_j = 0$ (with $\gamma_j \neq 0$ since we must suppose that the plane of no face is perpendicular to the xy-plane) and $a_j = \alpha_j/\gamma_j$, $b_j = \beta_j/\gamma_j$ and $d_j = \delta_j/\gamma_j$.

The coordinates (x_i, y_i) of the projected pictures of the vertices are known and the values z_i must be prescribed so that the quantities a_j, b_j, d_j be uniquely determined.

3. Examples

Example 3.1. The describing system of equations for the network of fig. 3.1 is

$$\begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 0 & 1 \\ -1 & 0 & 0 & R_2 & 0 \\ 0 & -1 & 0 & 0 & R_3 \end{pmatrix} \begin{pmatrix} u_2 \\ u_3 \\ i_1 \\ i_2 \\ i_3 \end{pmatrix} = \begin{pmatrix} u_1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

and the network is uniquely solvable if and only if $\det W = R_2 + R_3 \neq 0$.

Example 3.2. The describing system of equations for the first framework of fig. 1.2 is

$$\begin{pmatrix} x_1 - x_2 & x_2 - x_1 & 0 & y_1 - y_2 & y_2 - y_1 & 0 \\ x_1 - x_3 & 0 & x_3 - x_1 & y_1 - y_3 & 0 & y_3 - y_1 \\ 0 & x_2 - x_3 & x_3 - x_2 & 0 & y_2 - y_3 & y_3 - y_2 \end{pmatrix} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{y}_1 \\ \dot{y}_2 \\ \dot{y}_3 \end{bmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

and the framework is rigid if and only if $r(W) = 3$. One can readily verify that this holds if and only if the three joints are noncollinear (recall that collinearity means

$$\det\begin{pmatrix} x_i - x_j & y_i - y_j \\ x_i - x_k & y_i - y_k \end{pmatrix} = 0 \quad \text{for } i, j, k \in \{1, 2, 3\},\ i \neq j \neq k \neq i) .$$

Example 3.3. Consider the drawing of fig. 3.2. If the coordinates z_1, z_2, z_3, z_4 and z_5 are prescribed, they determine a polyhedron (a pyramid) with the given

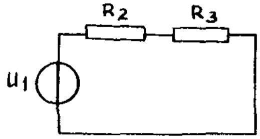


Figure 3.1.

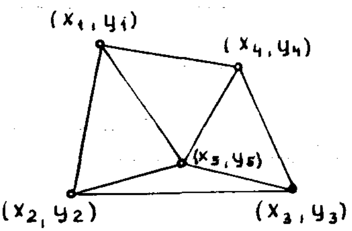


Figure 3.2.

projection if and only if

$$\det\begin{pmatrix} x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \\ x_4 & y_4 & z_4 & 1 \end{pmatrix} = 0 \quad \text{and} \quad \det\begin{pmatrix} x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \\ x_5 & y_5 & z_5 & 1 \end{pmatrix} \neq 0$$

since the vertices 1, 2, 3 and 4 must be coplanar and vertex 5 must not be on this plane.

4. Algebraic or combinatorial solutions?

The examples of the previous section illustrate that, at least theoretically, the problems have purely algebraic solutions. However, in the case of large scale systems, the size of the matrices will be large, hence round-off errors can arise during the numerical calculation. Such a "small" numerical mistake can mean that a determinant does not vanish.

Since our problems are qualitative (N is solvable or not, F is rigid or not, D is reconstructible or not), such errors are unacceptable. Hence our aim is to give solutions which are essentially combinatorical (with 0–1 arithmetic, free of round-off errors) and use as little numerical technique as possible.

In order to do this, we must distinguish between those singularities (unsolvability, nonrigidity, nonreconstructability) which are caused by the structure of the systems alone, and those which are caused together by the structure and the numerical parameters of the systems.

These distinctions will now be illustrated by a number of examples.

Example 4.1. The network of fig. 4.1 is unsolvable. This is caused by its structure only, no matter what the numerical values of R_1, R_5, R_7 and α are.

Example 4.2. The network of fig. 4.2 is unsolvable if and only if $\beta = 1 + R_1 \cdot R_4^{-1}$ holds.

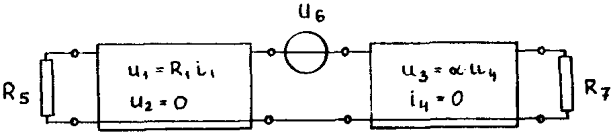


Figure 4.1.

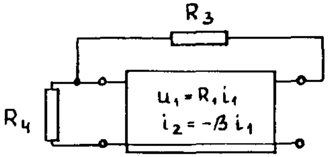


Figure 4.2.

Example 4.3. The planar framework of fig. 4.3 is nonrigid, independently of the coordinates of the six joints.

Example 4.4. The planar framework of fig. 4.4 is nonrigid if and only if the six joints are on a conic section.

Example 4.5. The drawing of fig. 4.5 cannot arise as the projection of a polyhedron.

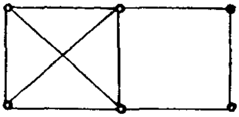


Figure 4.3.

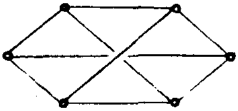


Figure 4.4.

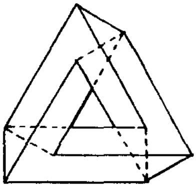


Figure 4.5.

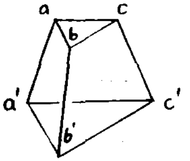


Figure 4.6.

Example 4.6. The drawing of fig. 4.6 is realizable if and only if the lines *aa'*, *bb'* and *cc'* intersect in a single common point.

These examples illustrate that a singularity may be purely structural (Examples 4.1, 4.3 and 4.5) or may depend on some additional algebraic relations among the numerical values of the parameters (Examples 4.2, 4.4 and 4.6). If we postulate that no such algebraic relationship exists (the so-called *genericity assumption*) then these problems become purely combinatorial.

This assumption may be justified in many practical situations. For example, parameters of electric devices depend on technological processes (e.g., two different resistors cannot be exactly equal or their parameters cannot satisfy any other a priori given algebraic relation either). The coordinates of a vertex in a drawing can also be considered as approximate (they are given, say, by a light pen on a graphical display in a computer aided design). However, one might have doubts about this approach as well. For example, there are electric devices (like transformers, gyrators) where the parameters may be considered as approximate but these approximate values appear in several equations and can cancel each other out. An additional problem is that the definition should possibly refer to the devices and not to one of the possible equivalent numerical descriptions of them. Hence the interested reader is referred for further comments to Murota and Iri (1985), Recski and Iri (1980), Recski (1989) and Sugihara (1986).

5. Solutions of the problems under the genericity assumption

Consider the column space matroid A of the coefficient matrix of the multiport equations (i.e., the vectorial matroid (see chapter 9) determined by the columns of the matrix); and the direct sum G of the cycle and the cocycle matroids of the network graph (see chapter 9).

Theorem 5.1. *The network is uniquely solvable under the genericity assumption if and only if the sum $A \vee G$ of these matroids is the free matroid.*

Proof (*sketch*). If A_1, A_2 and A_3 are the column space matroids of the matrices $\boldsymbol{A}_1$, $\boldsymbol{A}_2$ and $\boldsymbol{A}_3 = \binom{\boldsymbol{A}_1}{\boldsymbol{A}_2}$, respectively, then every independent subset of A_3 is independent in $A_1 \vee A_2$ as well (see Edmonds 1967). Hence our matroidal condition is always necessary. The reverse statement of Edmonds' theorem is usually not true. For example put $A_1 = (1 \quad 2)$, $A_2 = (2 \quad 4)$ and $A_3 = \left(\begin{smallmatrix}1 & 2\\ 2 & 4\end{smallmatrix}\right)$. Then $A_1 \vee A_2$ is the free matroid but A_3 is not, due to the cancellation $1 \cdot 4 - 2 \cdot 2 = 0$. Such cancellations, however, are excluded by the genericity assumption. □

Obviously, one can check this property by the matroid partition algorithm, see chapter 11.

Consider the graph G of the framework F, with n nodes and m edges. The definition of rigidity clearly implies $m \geqslant 2n - 3$ for planar frameworks. For brevity, we characterize *minimally rigid* planar frameworks only (where $m = 2n - 3$). The general characterization can be found in Lovász and Yemini (1982).

Theorem 5.2. *The planar framework is minimally rigid under the genericity assumption if and only if $m = 2n - 3$ and doubling any edge of G, the resulting graph can be covered by two spanning forests.*

Proof (*sketch*). The framework of fig. 4.3 shows that $m = 2n - 3$ cannot be sufficient: a condition $m' \leqslant 2n' - 3$ for every subgraph G' of G with n' nodes and m' edges is also necessary, to avoid local overbracing. A theorem of Laman (1970) states that $m = 2n - 3$ and this additional condition for every subgraph implies planar rigidity under the genericity assumption. This gives $m' \leqslant 2n' - 2$ for the graph G_e (obtained by doubling any edge e of G), which is just the condition of Theorem 10.5 of chapter 9 for the cycle matroid of G_e. □

Let F denote the set of faces of a drawing D. For every subset $X \subseteq F$ denote by $i(X)$ the number of incidences of the X-faces (i.e., 3 for triangles, 4 for quadrangles etc.) and by $v(X)$ the number of vertices incident to these X-faces.

Theorem 5.3. *The drawing is a projection of a convex polyhedron under the genericity assumption if and only if $i(X) \leqslant 3|X| + v(X) - 4$ for every $X \subseteq F$, $|X| \geqslant 2$.*

This last theorem (see Sugihara 1984 for a proof) essentially asks for the minimization of the submodular function $3|X| + v(X) - i(X)$. Due to the special nature of the problem it can be reduced to the constructions of matchings in bipartite graphs, hence its polynomial time-complexity is straightforward, without any reference to Grötschel et al. (1981). As we indicated, the other two theorems lead to matroid partition algorithms (Edmonds 1968), hence their time-complexity is also polynomial.

6. Various remarks

In this last section we wish to give a few hints about the history of the problems, about more general results and open problems.

The idea of characterizing unique solvability of electric networks by a combinatorical condition is essentially due to Kirchoff (1847): If a network consists of voltage and current sources and Ohmic resistors (i.e., 1-ports) only then the condition of Theorem 5.1 reduces to the existence of a spanning forest of the network graph which contains all the voltage source edges and none of the current source edges. Kirchhoff proved that this condition is sufficient even if the genericity assumption is not postulated but all resistors are positive.

Combinatorial methods for multiport networks are used since (Coates (1958) and Mayeda (1958) while minimax relations and essentially matroidal tools date back to Iri (1968), Kishi and Kajitani (1968), Ohtsuki et al. (1968), Ozawa (1975). The idea of using the matroid sum is independently due to Iri and Tomizawa (1975), Narayanan (1974) and Recski (1973).

Our definition of multiports in section 1 was very restrictive. Network theorists would call our multiports linear homogeneous and memoryless. However, many combinatorial results can be extended for wider classes of networks as well.

The application of combinatorics in statics dates back to at least Maxwell (1864a), Bow (1873) and Cremona (1879); the graphical method they devised to determine the stresses in the bars (the so-called Cremona diagrams) is essentially an application of the duality of planar graphs. The conditions $m = 2n - 3$ and $m' \leqslant 2n' - 3$ for every subgraph has been known to be necessary (and not sufficient) for minimal planar rigidity for more than 100 years but Laman's (1970) theorem seems to be the first good characterization for minimal planar rigidity under the genericity assumption. Theorem 5.2 which gives a polynomial algorithm as well, is due to Lovász and Yemini (1982).

The analogous necessary condition for the 3-dimensional space is $m = 3n - 6$ and $m' \leqslant 3n' - 6$ for the subgraphs. However, this is not sufficient under the genericity assumption, see fig. 6.1 (Asimow and Roth 1979). A good characterization for minimal 3-dimensional rigidity is an outstanding open problem of "structural topology".

Theorem 5.3 is due to Sugihara (1986) and the reader is referred to his book for many related results. However, one should point out that there is a deeper geometric connection between Problems 1.2 and 1.3, see Maxwell (1864b).

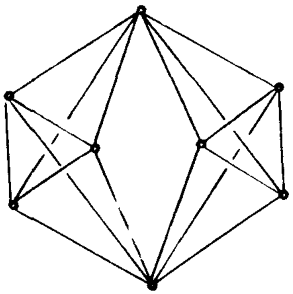


Figure 6.1.

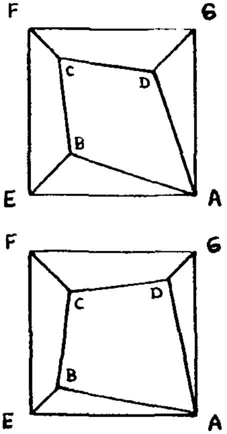


Figure 6.2.

Theorem 6.1. *A planar framework with n joints and m* = 2*n* − 3 *rods is rigid if and only if no part of it is the projection of a* 3*-dimensional polyhedron.*

For example, the second framework of fig. 6.2 is rigid, the first one is not. This can be seen by denoting the intersection of lines *BC* and *EF* by *H* and that of *CD* and *FG* by *I* and checking whether *A*, *H* and *I* are collinear (see fig. 6.3).

These and related models have actually been used by engineering communities. For example, a network analysis program based on Theorem 5.1 has been developed at the Technical University of Denmark, Lyngby (see Petersen 1979); a program to reconstruct polyhedra from projected images based on Theorem 5.3, at the Electrotechnical Laboratory in Tsukuba, Japan (see Sugihara 1984); while a

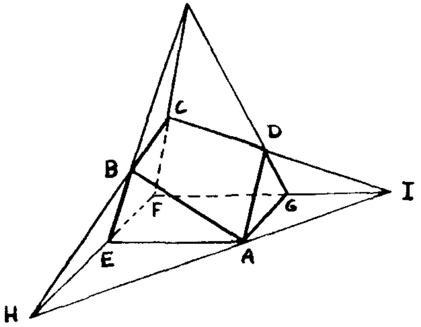


Figure 6.3.

dynamic process simulation system, using strongly related mathematical tools, is actually used in some ten chemical companies in Japan (see chapter 2 in Murota 1987).

References

Asimow, L., and B. Roth

[1978–79] The rigidity of graphs, Part I, *Trans. Amer. Math. Soc.* **245**, 279–289; Part II, *J. Math. Anal. Appl.* **68**, 171–190.

Bow, R.M.

[1873] *On the Economics of Construction in Relation to Framed Structures* (E. & F.N. Spon, London).

Coates, C.L.

[1958] General topological formulas for linear network functions, *IRE Trans. Circuit Theory* **CT-5**, 42–54.

Crapo, H.H., and W. Whiteley

[1994] *Geometry of Rigid Structures* (Cambridge University Press, Cambridge) to appear.

Cremona, L.

[1879] *Le Figure Reciproche nella Statica Grafica* (Milano). English translation: 1890, *Graphical Statics* (Oxford University Press, London).

Edmonds, J.R.

[1967] Systems of distinct representatives and linear algebra, *J. Res. Nat. Bureau Standards B* **71**, 241 245.

[1968] Matroid partition, in: *Mathematics of the Decision Sciences I*, eds. G.B. Dantzig and A.F. Veinott, *Lectures in Applied Mathematics*, Vol. 11 (AMS, Providence, RI) pp. 335–345.

Grötschel, M., L. Lovász and A. Schrijver

[1981] The ellipsoid method and its consequences in combinatorial optimization, *Combinatorica* **1**, 169–197.

Iri, M.

[1968] A min–max theorem for the ranks and term-ranks of a class of matrices: An algebraic approach to the problem of the topological degrees of freedom of a network, *Trans. Inst. Electron. Comm. Eng. Japan A* **51**, 180–187.

[1983] Applications of matroid theory, in: *Mathematical Programming: The State of the Art*, eds. A. Bachem, M. Grötschel and B. Korte (Springer, Berlin) pp. 158–201.

Iri, M., and S. Fujishige

[1981] Use of matroid theory in operations research, circuits and systems theory, *Int. J. Syst. Sci.* **12**, 27–54.

Iri, M., and N. Tomizawa

[1975] A unifying approach to fundamental problems in network theory by means of matroids, *Trans. Inst. Electron. Comm. Eng. Japan A* **58**, 33 40.

Kirchhoff, G.

[1847] Ueber die Auflösung der Gleichungen, auf welche man bei der Untersuchungen der linearen Vertheilung galvanischer Ströme geführt wird, *Ann. Phys. Chemie* **72**, 497 508.

Kishi, G., and Y. Kajitani

[1968] Maximally distinct trees in a linear graph, *Trans. Inst. Electron. Comm. Eng. Japan A* **51**, 196 203.

Laman, G.

[1970] On graphs and rigidity of plane skeletal structures, *J. Eng. Math.* 4, 331–340.

Lovász, L., and Y. Yemini

[1982] On generic rigidity in the plane, *SIAM J. Algebra Discrete Methods* 3, 91–98.

Maxwell, J.C.

[1864a] On reciprocal figures and diagrams of forces, *Philos. Mag. (4)* **27**, 250–261.

[1864b] On the calculation of the equilibrum and stiffness of frames, *Philos. Mag. (4)* **27**, 294 299.

Mayeda, W.

[1958] *Topological formulas for active networks*, Int. Techn. Report 8 (DA-11–022-ORD-1983, University of Illinois).

Murota, K.
[1987] *Systems Analysis by Graphs and Matroids: Structural Solvability and Controllability* (Springer, Berlin).

Murota, K., and M. Iri
[1985] Structural solvability of systems of equations – a mathematical formulation for distinguishing accurate and inaccurate numbers in structural analysis of systems, *Japan J. Appl. Math.* **2**, 247–271.

Narayanan, H.
[1974] *Theory of matroids and network analysis*, Ph.D. Thesis (Indian Institute of Technology, Bombay).

Ohtsuki, T., Y. Ishizaki and H. Watanabe
[1968] Network analysis and topological degrees of freedom, *Trans. Inst. Electron. Comm. Eng. Japan A* **51**, 238–245.

Ozawa, T.
[1975] Solvability of linear electric networks, *Mem. Fac. Eng. Kyoto Univ.* **37**, 299–315.

Petersen, B.
[1979] Investigating solvability and complexity of linear active networks by means of matroids, *IEEE Trans. Circuits Systems* **CAS-26**, 330–342.

Recski, A.
[1973] On partitional matroids with applications, *Colloq. Math. Soc. János Bolyai* **10**, 1169–1179.
[1984] Applications of combinatorics to statics – a survey, *Rend. Circ. Mat. Palermo II* **3**, 237–247.
[1989] *Matroid Theory and its Applications in Electric Network Theory and in Statics* (Springer/Akadémiai Kiadó, Berlin/Budapest).

Recski, A., and M. Iri
[1980] Network theory and transversal matroids, *Discrete Appl. Math.* **2**, 311–326.

Roth, B.
[1981] Rigid and flexible frameworks, *Amer. Math. Monthly* **88**, 6–21.

Spriggs, J.R., and R.A. Snay
[1982] An algorithm for testing the solvability of horizontal networks, in: *Proc. 25th Spring Meeting, Philadelphia, PA* (American Geophysical Union, Washington, DC).

Sugihara, K.
[1984] An algebraic and combinatorial approach to the analysis of line drawings of polyhedra, *Discrete Appl. Math.* **9**, 77–104.
[1986] *The Machine Interpretation of Line Drawings* (MIT Press, Cambridge, MA).

CHAPTER 37

Combinatorics in Statistical Physics

C.D. GODSIL

Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ont. N2L 3G1, Canada

M. GRÖTSCHEL

Konrad-Zuse-Zentrum für Informationstechnik, Heilbronner Str. 10, D-10711 Berlin-Wilmersdorf, Germany

D.J.A. WELSH*

Merton College and Mathematical Institute, University of Oxford, Oxford OX1 4JD, UK

Contents

* Supported by SFB 303 (DFG).

HANDBOOK OF COMBINATORICS
Edited by R. Graham, M. Grötschel and L. Lovász

1. Introduction

Combinatorics and physics interact in various ways. It is impossible to survey here all the connections. We concentrate in this chapter on statistical physics since several of the most basic problems in this area have a combinatorial flavour.

Sections 2 and 3 of this chapter are concerned with two of the most fundamental areas, namely the study of the Ising model and the theory of percolation processes. Both of these areas of research are huge, but they share the common property that some of the most primitive and easily stated problems are, after more than thirty years of research, still largely unanswered.

In section 4 we present, among other models, some of the classical enumeration problems of statistical physics; again there are many open questions and very few exact results.

Sections 5 and 6 are concerned with two of the (relatively few) "techniques" which have been developed to deal with the sort of problems we are discussing. Transfer matrices and subadditive function theory are basic tools in this area of mathematical physics. This is illustrated by a simplified version of the dimer problem, it amounts to counting the number of ways of placing dominoes on a rectangular chessboard.

Finally we illustrate in Section 7 the application of ideas from combinatorial optimization to statistical physics by showing how the problem of finding the ground states of a spin glass model may be reduced to a very basic, though difficult (NP-hard) problem in discrete optimization.

2. The Ising model

The density of water varies as a function of temperature, and generally as a continuous function. Of course the variation is not continuous in the neighbourhood of the boiling point, nor at the freezing point. Although we are accustomed to such behaviour, it is paradoxical. The forces acting between the individual molecules vary continuously as the temperature varies. Why then should there be a change of state at certain temperatures? Statistical physics is devoted to the attempt to understand this behaviour.

As is customary in science and mathematics, the study begins by setting up a grossly simplified model. We assume that the system consists of a finite number of particles, and that the system is at any instant in one of a number of states. The behaviour of the system is governed by its Hamiltonian H, which is a function of the state. Its value $H(\sigma)$ is equal to the energy of the system in state σ. Examples of Hamiltonians will be given later in this section and also in section 7. The *partition function* of the system is defined to be

$$Z = Z(T) = \sum_{\sigma} \exp[-H(\sigma)/(kT)]\,. \tag{2.1}$$

Here T is the temperature of the system and k is Boltzmann's constant. If the system consists of N particles, we sometimes write Z_N in place of Z. It is taken as an axiom that all large-scale properties of the system are determined by Z. (Sometimes an attempt is made to disguise the fact that this is an axiom, and not a theorem. Also a system can have more than one partition function; the one we have just defined is the *canonical* partition function.)

In stochastic versions of the problem it is assumed that in the stationary state the probability that the system is in the state σ is given by

$$Z^{-1}\exp[-H(\sigma)/(kT)]$$

and that the *free energy* of the system is

$$F = -kT\log Z\,.$$

The latter is a particularly important parameter of the system, and explains the fact that one deals with $\log Z$ as often as with Z. We are interested in the behaviour of Z for systems with a large number of particles, as the temperature T ranges over the positive reals. The value of Z depends on the number of particles N in the system as well as on the temperature. In all cases of interest to us, $\log Z$ is a linear function of N when all other parameters are fixed. The number of particles in any realistic physical system is, for all mathematical purposes, infinite. Hence we are led to study

$$\lim_{N\to\infty}\frac{1}{N}\log Z_N(T)$$

as a function of T. Following Baxter (1982, p. 14), we say that a model has been solved if its free energy is known. The phase transitions of the model correspond to the points, called *critical points*, at which the free energy is not an analytic function.

We now consider a typical and important system, the *Ising model*. We are given a graph $G=(V,E)$ embedded in $\mathbb{R}^2$ or $\mathbb{R}^3$. There is an atom placed at each vertex. Each atom has a spin associated to it, this spin takes only two values. The energy of the system is understood to be the sum of the energies due to the interaction of each pair of atoms. The contribution due to a pair of atoms will be assumed to depend only on whether they are adjacent in G or not. The interaction is completely determined by whether the given pair have the same spin, or not.

The state of the system can be represented by a function σ from $V(G)$ into the set $\{-1,1\}$ and the Hamiltonian $H(\sigma)$ will be a sum over the edge set E of G. Writing σ_i for the state of atom i, we find that the partition function of this system at temperature T is

$$Z(G)=\sum_{\sigma}\exp\Big[-\sum_{ij\in E}\beta\sigma_i\sigma_i\Big]\,. \tag{2.2}$$

The constant $\beta = J/(kT)$ will vary inversely with the temperature and is proportional to the interaction J.

The graphs in which physicists are interested are usually infinite. However, they are usually the limit, in a natural sense, of a sequence of finite graphs. This will be made clearer by the examples which follow. The most important cases of the Ising model are when G is either the 2-dimensional square lattice, or the 3-dimensional cubic lattice. The solution of the 2-dimensional Ising problem on the square lattice was a major achievement of Onsager in 1944. (For an account of this, and any other historical remarks in this section, see Thompson 1972.) The 3-dimensional model is still unsolved.

There are some important extensions of the Ising model. We assumed implicitly that each of the two states available to an individual atom was equally likely. However, if there is an external magnetic field acting then one of the two states becomes more probable. The 2-dimensional Ising model has only been solved under the assumption that there is no external field. (The presence of an external field in any model is a major complication.) Another possibility is that the interactions between a pair of adjacent atoms may not be independent of the edge. (This is certainly a physically reasonable possibility.) Thus on the square lattice, the interactions arIsing from the horizontal edges may differ from the interactions on the vertical edges. Allowing for this does not usually cause problems; on the contrary it can even be useful, as we will see.

A question which may well have arisen by now is, what does all this have to do with combinatorics? To explain this, we study the basic Ising model on the square lattice. Let G_n denote the Cartesian product $P_n \times P_n$ of two paths with n vertices. Thus G_n has n^2 vertices and, for large n, may be viewed as an approximation to the infinite square lattice. By expanding the exponential in (2.2) and since $\sigma_i\sigma_j$ takes only the values $+1$ and -1, we have

$$\exp(\beta\sigma_i\sigma_j) = \cosh(\beta) + \sigma_i\sigma_j \sinh(\beta) = \cosh(\beta)(1 + \sigma_i\sigma_j \tanh(\beta)),$$

whence the partition function for G_n at temperature T can be expressed as

$$Z(G_n) = \sum_{\sigma} (\cosh(\beta))^{|E(G_n)|} \prod_{ij\in E(G_n)} (1 + \sigma_i\sigma_j \tanh(\beta)) .$$

With some ingenuity (see, e.g., section 6.1 of Thompson 1972 or Biggs 1977, p. 22) this may be rewritten as

$$Z(G_n) = 2^{|V(G_n)|}(\cosh(\beta))^{|E(G_n)|} \sum_{l\geq 0} N(l)(\tanh(\beta))^l , \tag{2.3}$$

where $N(l)$ is the number of spanning subgraphs of G_n with l edges and all vertices having even valency. (These are called the Eulerian subgraphs of G_n.) This shows that determining the partition function for the Ising model is equivalent to the purely combinatorial problem of enumerating the Eulerian subgraphs of G_n.

It should be noted that (2.3) is valid with any graph G in place of G_n. In particular if we replace G_n by P_n then we obtain

$$Z(P_n) = 2^n(\cosh(\beta))^{n-1} .$$

From this we can deduce that

$$\lim_{n\to\infty} (Z(P_n))^{1/n} = 2\cosh(\beta)\,.$$

Since $\cosh(\beta)$ is an analytic function, it follows that the Ising model on the infinite path does not have a critical point. As we noted earlier in this section, Onsager showed that the Ising model on the infinite square lattice does have a phase transition. (See chapter 5 of Thompson 1972.)

A short introduction to the Ising problem will be found in Cipra (1987).

Partition functions and rank polynomials

We now show how Fortuin and Kasteleyn (1972) demonstrated that the Ising and other physical problems could be related to the Whitney rank polynomial or Tutte polynomial (see chapter 9 by Welsh).

Let G be a graph, which now may have loops and multiple edges. Any subset S of $E(G)$ forms a spanning subgraph of G, with the same vertex set as G, and edge set S. The *rank* of S is defined to be $|V(G)|$, less the number of connected components in the subgraph formed by S. We will denote it by $r(S)$. The *rank polynomial* of a graph G is defined to be

$$R(G; x, y) = \sum_{S\subseteq E(G)} x^{r(E)-r(S)} y^{|S|-r(S)}\,.$$

The rank polynomial has some interesting properties. If G is the disjoint union of graphs G_1 and G_2 then

$$R(G; x, y) = R(G_1; x, y) R(G_2; x, y)\,. \tag{2.4}$$

If $e \in E(G)$, let $G\backslash e$ be the graph obtained by deleting e from G, and let G/e be the graph obtained by contracting e (i.e., by deleting e and then identifying its end points). Then, if e is not a cut-edge or a loop, one can show that

$$\begin{aligned} R(G; x, y) &= \sum_{S\subseteq E(G), e\in S} x^{r(S)} y^{|S|-r(S)} + \sum_{S\subseteq E(G), e\notin S} x^{r(S)} y^{|S|-r(S)}\\ &= R(G/e; x, y) + R(G\backslash e; x, y)\,. \end{aligned} \tag{2.5}$$

In the remaining cases we have

$$R(G; x, y) = \begin{cases} (1+x)R(G/e; x, y)\,, & \text{if } e \text{ is a cut-edge}\,,\\ (1+y)R(G\backslash e; x, y)\,, & \text{if } e \text{ is a loop}\,. \end{cases}$$

Now consider the partition function for the Ising model on a graph G, which can be written in the form

$$Z(G) = \sum_{\sigma} \prod_{ij\in E(G)} \lambda^{\sigma_i\sigma_j}\,,$$

where $\lambda = \exp\beta$. The product $\sigma_i\sigma_j$ is either 1 or -1. Define E_σ^+ to be the set of edges ij of G such that $\sigma_i\sigma_j = 1$ and let E_σ^- be the remaining edges of G. Let

$m=|E(G)|$. Then we have

$$Z(G)=\sum_{\sigma}\lambda^{|E_\sigma^+|-|E_\sigma^-|}=\sum_{\sigma}\lambda^{m-2|E_\sigma^-|}\,.$$

If $e=ij$ is a fixed edge in $E(G)$, not a loop or a cut-edge, it follows that

$$\begin{aligned}Z(G)&=\sum_{\sigma_i=\sigma_j}\lambda^{m-2|E_\sigma^-|}+\sum_{\sigma_i\neq\sigma_j}\lambda^{m-2|E_\sigma^-|}\\&=\lambda Z(G/e)+\lambda^{-1}(Z(G\backslash e)-Z(G/e))\\&=(\lambda-\lambda^{-1})Z(G/e)+\lambda^{-1}Z(G\backslash e)\,.\end{aligned}\tag{2.6}$$

We can now use the following theorem of Oxley and Welsh (1979).

Theorem 2.7. *Let f be a real-valued function defined on graphs which satisfies the recursion*

$$f(G)=af(G/e)+bf(G\backslash e)$$

when e is an edge of G and not a loop or cut-edge, and

$$f(G)=\begin{cases}(1+x)f(G/e)\,, & \text{if } e \text{ is a cut-edge}\,,\\(1+y)f(G\backslash e)\,, & \text{if } e \text{ is a loop}\,,\end{cases}$$

where $1+x$ and $1+y$ are the values taken by f on a cut-edge and loop respectively. Then if G has n vertices, m edges and c components, we have

$$f(G)=b^{m-n+c}a^{n-c}R\Big(G;\frac{1+x}{a}-1,\frac{1+y}{b}-1\Big)\,.$$

It follows that the partition function for the Ising model on a graph G is determined by its rank polynomial. From (2.6) we see that we can apply the previous theorem with $Z(G)$ in place of $f(G)$. Then

$$a=\lambda-\lambda^{-1}\,,\qquad b=\lambda^{-1}$$

and

$$1+x=\lambda+\lambda^{-1}\,,\qquad 1+y=\lambda^{-1}\,.$$

Theorem (2.7) now yields that

$$Z(G)=\lambda^{-m}(\lambda^2-1)^{n-c}R\Big(G;\frac{2}{\lambda^2-1},\lambda^2-1\Big)\,.$$

A natural extension of the Ising model is to allow the spins to take more than two values. More precisely, if we allow the spin at each vertex to take values from the set $\{1,2,\ldots,q\}$ and then define the partition function Z by $Z=\Sigma_\sigma\exp[\Sigma_{ij\in E}\,\beta\delta(\sigma_i,\sigma_j)]$ where δ is the usual delta function, we have what is known as the *q-state Potts model*.

Using a similar argument to that just given it is easy to see that again Z satisfies a contraction–deletion recurrence formula. Hence for any graph G, Z is an evaluation of the rank polynomial of G; though along a different curve in the xy-plane, namely $xy = q$. For a proof of this and for details of the way in which the percolation and ice models to be discussed below can be represented in terms of the rank polynomial see Welsh (1990) or the original paper of Fortuin and Kasteleyn (1972).

For excellent rigorous mathematical treatments of these topics we refer to the monographs of Ruelle (1969) or Thompson (1972).

3. Percolation processes

As its name suggests, percolation theory is concerned with flow in random media. Its origin in the work of Broadbent and Hammersley (1957) was as a model for molecules penetrating a porous solid, electrons migrating over an atomic lattice, a solute diffusing through a solvent, or a disease infecting a community.

As an example of percolation in the wider sense consider the following problem in communication theory.

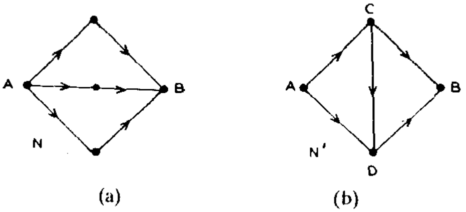
Example (*Random graphs and reliability*). Let N be the network shown in fig. 3.1(a). Suppose each directed edge has probability p of being reliable, that is, allowing a message to pass. Suppose further that the reliability of each edge is independent of the reliability of any other edge. What is the probability that there is a path from A to B consisting only of reliable edges?

Denoting this event by $A \sim B$, simple calculation shows that it is just the probability that not all the routes from A to B are unreliable. Since the routes have no edge in common we are dealing with independent random variables and we have

$$\Pr[A \sim B] = 1 - (1 - p^2)^3 .$$

However, if we try the same problem for the network N', of fig. 3.1(b) the problem becomes much more complicated. This is due solely to the *dependence* in N' of the events "the route $ACDB$ is reliable" and "the route ACB is reliable".

This problem illustrates the intrinsic difficulty of percolation problems – stochastic dependence occurs in all but the most trivial cases and makes computation



(a) (b)

Figure 3.1.

very difficult. Indeed, even with the speed of modern computing machines it is still impractical to determine the reliability of moderate-sized networks. In the language of computational complexity the problem is #P-hard (see chapter 29 by Shmoys and Tardos).

In *classical percolation theory* we are concerned with the probability of infinite clusters in a "regular crystal lattice". The definition of what exactly is meant by a "regular crystal lattice" is rather difficult to formulate precisely – it varies from author to author. For the purposes of this chapter it can be regarded as typified by the regular lattices shown in fig. 3.2, though of course the physically most interesting cases are when the lattice is 3-dimensional.

Bond percolation

Suppose now that we fix attention on the 2-dimensional square lattice, and suppose that there is a supply of fluid at the origin and that each edge of $\mathscr{L}_2$ allows fluid to pass along it with probability p, independently for each edge. Let $P_n(p)$ be the probability that fluid spreads to at least n vertices. Thus

$$P_1(p) = 1 ,$$

$$P_2(p) = 1 - (1 - p)^4 ,$$

and in theory $P_N(p)$ can be calculated for any integer N. However, the reader will rapidly find it prohibitively time consuming. The case $N = 7$ is a fair piece of work! Obviously,

$$P_N(p) \geqslant P_{N+1}(p)$$

and hence the limit

$$P(p) = \lim_{N\to\infty} P_N(p)$$

exists and represents the probability that fluid spreads an infinite distance from the origin.

Very little has been rigorously proved about $P(p)$. For example, even though $P_N(p)$ is a polynomial in p and hence we would expect $P(p)$ to be a continuous function of p, this has not yet been proved. It is clear that there exists a *critical*

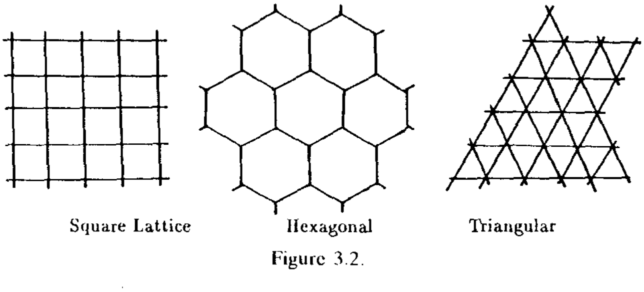


Figure 3.2.

probability p_c such that

$$p < p_c \Rightarrow P(p) = 0 \quad \text{and} \quad p > p_c \Rightarrow P(p) > 0 .$$

However, determining the value of this critical probability is, as we will see, a very difficult problem. Monte Carlo simulations suggest that for all well-known lattices the behaviour of $P(p)$ has roughly the same S-shaped form as shown in fig. 3.3.

Atom or site percolation

In atom percolation, instead of each edge being randomly blocked with probability $1-p$ or open with probability p, each vertex is blocked independently with probability $1-p$ or open with probability p. Again we are interested in the probability of fluid spreading locally or an infinite distance.

Exactly analogous results hold for atom percolation as for bond percolation, though of course the numerical values of the critical probabilities p_c and percolation probabilities $P(p)$ differ. In one sense atom percolation is the more important since any bond percolation problem on a lattice $\mathscr{L}$ can be turned into an atom percolation problem on a related lattice $\hat{\mathscr{L}}$, namely the *line graph* of $\mathscr{L}$.

One of the few relatively easy results which has been proved is the following due to Fisher (1961) and Hammersley (1961a). For any regular lattice, if $P^{A}(p)$, $P^{B}(p)$ represent respectively the atom and bond percolation probabilities on the lattice then

$$P^{A}(p) \leq P^{B}(p) , \quad 0 \leq p \leq 1 .$$

Clearly this implies that for any lattice the critical probability for atom percolation is at least as big as the critical probability for bond percolation.

The cluster problem

An alternative approach to percolation theory is the study of the distribution of white and black clusters when the edges (or vertices) of a graph are independently painted white with probability p and black with probability $q = 1 - p$.

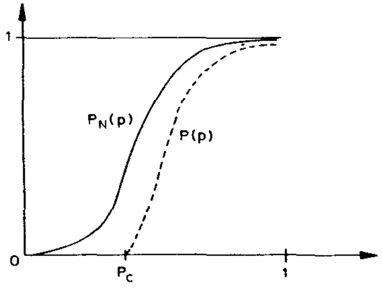


Figure 3.3.

Again we shall concentrate on the edge problem for the square lattice. A white *cluster* is a maximal connected subset of white edges of the lattice. The two main quantities of physical interest are:

(a) the average number of white clusters;

(b) the average number of vertices in a white cluster.

To be more precise let $\mathscr{L}_m$ denote a square section of the square lattice containing m^2 vertices and hence $2(m-1)^2$ edges. If ω denotes a particular black/white painting of $\mathscr{L}_m$ then let $c_m(\omega)$ denote the number of white clusters and let its average value over all paintings ω be denoted by $K_m(p)$.

Similarly if we let the distinct clusters under ω be labelled $A_1, \ldots, A_{c(\omega)}$, we define

$$S_m(p) = \left\langle \left[\frac{|V(A_1)| + \cdots + |V(A_{c(\omega)})|}{c_m(\omega)} \right] \right\rangle ,$$

where $|V(A_i)|$ denotes the number of vertices ion A_i, and $\langle \cdot \rangle$ denotes the expectation over all black and white paintings. Thus $S_m(p)$ is the average number of vertices in a white cluster.

Note that if isolated points are not counted as clusters then the expected number of clusters in this sense is given by $K_m(p) - m^2q^4$ where $q = 1 - p$. This is because the probability that a particular vertex forms an isolated cluster is just the probability that the four edges incident with it are painted black, that is q^4. Thus the average number of isolated points amongst the white clusters is m^2q^4.

The average number of black clusters is obviously $K_m(1-p)$ and the average number of vertices in a black cluster is obviously $S_m(1-p)$. Little more is known theoretically about either of these functions, other than that

$$K_m(p) \sim m^2\lambda(p) \quad \text{as } m \to \infty ,$$

where λ is an undetermined function of p.

Roughly speaking the quantities $K_m(p)$ and $S_m(p)$ are reciprocal, though theoretically all that has been proved is that

$$S_m(p) \geq m^2/K_m(p) .$$

For p greater than the critical probability p_c we have a positive probability of an infinite white cluster in $\mathscr{L}_\infty$. Hence, a fortiori, as $p \to p_c$ the average size of a cluster tends to ∞. Numerical evidence suggests that, as p approaches p_c from below, there exist constants C and γ such that as $m \to \infty$, $S_m(p) \to S(p)$ where

$$S(p) \sim C(p_c - p)^{-\gamma} ,$$

where, moreover, γ is an invariant depending only on the dimensionality of the lattice.

One of the most interesting theoretical results on the cluster problem is the following theorem of Harris (1960).

Theorem 3.1. *For the cluster problem on the infinite square lattice, if p is strictly greater than the critical probability then, with probability one, the set of white edges contains only one infinite component.*

Extensions of this to higher dimensions can be found in Grimmett (1989).

Determining the critical probability

The problem of finding critical probabilities for particular lattices, first posed in 1957, has received great attention, but is still proving to be a remarkably difficult problem. A vast amount of numerical estimation (based on Monte Carlo methods, Padé approximations and the like) has been carried out, so good numerical estimates exist for most of the 2- and 3-dimensional lattices.

Theoretically much less is known. A landmark in the study of critical probabilities was the paper by Sykes and Essam (1964) which, though unrigorous, gave convincing arguments for believing that for bond percolation on a planar lattice the critical probabilities were related by

$$p_c(\mathscr{L}) + p_c(\mathscr{L}^*) = 1 , \tag{3.2}$$

where $\mathscr{L}^*$ is the planar dual of $\mathscr{L}$.

An obvious consequence of this is the following.

Theorem 3.3. *For bond percolation on the square lattice, the critical probability is* $\frac{1}{2}$.

This result was finally proved by Kesten (1980) by a series of ingenious arguments which have led to a rigorous proof by Wierman (1981) of the following result, again first shown unrigorously by Sykes and Essam (1964).

Theorem 3.4. *For bond percolation on the 2-dimensional triangular lattice (T) and the hexagonal lattice (H),*

$$p_c(T) = 2\sin(\pi/18) ,$$
$$p_c(H) = 1 - 2\sin(\pi/18) .$$

However, all of these arguments are very much restricted to 2 dimensions. A fundamental and very difficult problem is the following.

lem 3.5. Determine the critical probability of bond or site percolation on the nsional cubic lattice.

2-dimensional planar lattices there are many open problems. For following is known from Toth (1985) and Zuev (1987).

Theorem 3.6. *The critical probability of site percolation on the square lattice is between* 0.5095 *and* 0.68189.

However, this is a very wide range and we pose the following.

Problem 3.7. What is the critical probability for site percolation on the square lattice?

First passage percolation

This originated in the paper of Hammersley and Welsh (1965) as a model for a "time dependent" percolation process. It contains ordinary percolation as a special case and in its most general sense can be regarded as a randomized version of the shortest route problem in graphs.

Consider the square lattice in which each edge, is, independently, assigned a non-negative random length drawn from a known probability distribution F.

Let t_n be the random first passage (shortest) path length from the origin to $(n, 0)$ in this lattice and let $\tau(n)$ be its expected value over all possible *states* (that is distribution of lengths). The fundamental observations are that for $m, n \in \mathbb{Z}$

$$\tau(m+n) \leq \tau(m) + (n) \tag{3.8}$$

so that by the theory of subadditive functions

$$\lim_{n\to\infty} \frac{\tau(n)}{n} = \inf_n \frac{\tau(n)}{n} = \mu \tag{3.9}$$

exists.

The *time constant* μ depends only on the distribution F and is, like the critical probability of ordinary percolation, a not very well-understood lattice invariant. For example when the lengths are uniformly distributed in $[0, 1]$ it is known from Monte Carlo studies that $\mu \simeq 0.323$, but its exact evaluation seems a hopelessly intractable problem.

Apart from its intrinsic interest, first passage percolation led Hammersley and Welsh (1965) to set up a theory of subadditive stochastic processes which are now a fundamental tool in probability and probabilistic combinatorics, see for example Kingman (1973).

Correlated percolation

In an ideal world one would like to be able to remove the restriction that the random component associated with an edge in each of the above models was independent of all other edges. This is the subject of correlated percolation which is now a topic of considerable interest in the physical literature but for which (understandably) there are very few theoretical results.

For comprehensive rigorous accounts of what is now a huge area of research in statistical physics we refer to the monographs of Smythe and Wierman (1978), Kesten (1982) and Grimmett (1989).

4. Enumeration and related problems

Several fundamental problems in statistical physics and related areas of the natural sciences reduce to enumerating structures of different types. In this section we discuss a few of the most studied and basic problems of this nature.

Animals or polyominoes

Consider the 2-dimensional square lattice $\mathscr{L}_2$ with origin 0. An *animal* or *polyomino* of *n cells* is a connected subgraph of $\mathscr{L}_2$ containing 0 and having n vertices. Let $a(n)$ denote the number of distinct animals having n cells. Then clearly $a(1) = 1$, $a(2) = 4$ and counting the 3-celled animal types illustrated in fig. 4.1, we see $a(3) = 18$.

The fundamental problem which is now at least 30 years old is to determine the form of $a(n)$ for the different lattices. However, as with percolation theory, rigorous exact results about animals are pretty scarce.

First we will point out the connection between animals and percolation theory. Suppose that we could determine $a(n, b)$, the number of animals with n cells and b boundary cells. (As the name suggests a cell is a *boundary* cell of a specific animal A if it is a vertex of $\mathscr{L}$ which is not in A but is adjacent to a vertex of A.) Then from $a(n, b)$ it is not difficult to compute the average cluster size in a percolation model. From this we get good bounds for the critical probability.

Other applications of animals are to growth problems and as models of branched polymers with excluded volume.

Now let us turn to some basic results about $a(n)$ for the square lattice. A straightforward counting argument gives

$$2^n \leqslant a(n) \leqslant (6.25)^n . \tag{4.1}$$

It is also easy to prove that, for any positive integers m, n,

$$a(m + n) \geqslant a(m)a(n) . \tag{4.2}$$

Proof: Each animal has a top right corner and a bottom left corner. By "sticking"

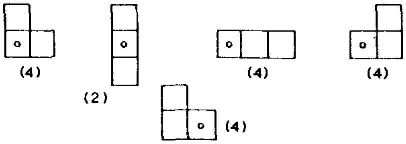


Figure 4.1.

the bottom left corner of an n-celled animal to the top right corner of an m-celled animal we obtain an $(m+n)$-celled animal.

By the basic property of subadditive functions, (4.1) and (4.2) give the fundamental result which holds (by analogous argument) for any regular lattice.

Theorem 4.3. *For any lattice $\mathscr{L}$ there exists a constant $a(\mathscr{L})$ such that if $a(n)$ denotes the number of n-celled animals of $\mathscr{L}$ then*

$$\lim_{n\to\infty} a(n)^{1/n} = \sup_{n\geq\infty}[a(n)]^{1/n} = a(\mathscr{L})\,.$$

Determining the limiting constant $a(\mathscr{L})$ exactly seems to be very difficult and even the best known bounds are not very tight. For example, in the most studied case of the square lattice, concatenation methods coupled with computer counts give the best known lower bound of 3.79 for $a(\mathscr{L}_2)$ while the best upper bound gives $a(\mathscr{L}_2)\leqslant 4.65$. There are reasons for believing that $a(\mathscr{L})$ is just above 4 in the case of this lattice.

For more details on these methods, the corresponding results for other lattices and a discussion of related problems we refer to a recent excellent review of Whittington and Soteros (1990).

Self-avoiding walks and polygons

Another counting problem closely connected with percolation theory and similar in spirit to the animal problem of the previous section is the following. A *self-avoiding walk* of length n on a lattice $\mathscr{L}$ is a path of n edges which has one endpoint at the origin. If $f(n)$ denotes the number of such self-avoiding walks, on the square lattice then clearly $f(1)=4$, $f(2)=12$ and in general it is easy to show that

$$2^n \leqslant f(n) \leqslant 4.3^{n-1}\,. \tag{4.4}$$

Using the submultiplicative property

$$f(m+n)\leqslant f(m)f(n)$$

and a similar bound to (4.4) for a general lattice, leads to the fundamental result.

Theorem 4.5. *For any lattice $\mathscr{L}$, there exists a constant $\mu=\mu(\mathscr{L})$ (known as the connective constant) such that if $f_{\mathscr{L}}(n)$ denotes the number of self-avoiding walks of length n on $\mathscr{L}$ then*

$$\lim_{n\to\infty} f_{\mathscr{L}}(n)^{1/n} = \inf_n [f_{\mathscr{L}}(n)]^{1/n} = \mu(\mathscr{L})\,.$$

Determining μ exactly for any lattice except the regular tree has been a much studied problem since it was first posed in 1957. Even good bounds seem to be difficult to obtain. For example, for the 2-dimensional square lattice, the best bounds so far known are $2.57\leqslant\mu\leqslant 2.73$.

A closely related quantity is $g_{\mathscr{L}}(n)$ which counts the number of self-avoiding polygons of n steps. Clearly $g_{\mathscr{L}}(n) \leqslant f_{\mathscr{L}}(n)$ but Hammersley (1961b) proved that for any lattice with connective constant $\mu(\mathscr{L})$,

$$\lim_{n\to\infty} g_{\mathscr{L}}(n)^{1/n} = \mu(\mathscr{L}) .$$

There are physical reasons (based on renormalisation arguments) and some numerical evidence which support the intriguing conjecture that there exist constants α and β such that

$$f_{\mathscr{L}}(n) \sim n^{\alpha}\mu^{n} , \qquad g_{\mathscr{L}}(n) \sim n^{\beta}\mu^{n}$$

and that α, β are dimensional invariants, in other words they only depend on the dimension of the lattice. Since Kesten (1963), there has been a great deal of rigorous mathematical progress notably by Hara and Slade (1992). For more details see Madras and Slade (1993).

The ice problem

As its name suggests the "ice problem" originates in the statistical physics associated with models used to calculate the residual entropy of "square ice". In its most general form an *ice model* specifies a set of allowable configurations at each vertex. All allowable configurations are of equal thermodynamic weight and the problem reduces to calculating the partition function, that is, enumerating the number of allowable configurations.

Probably the most studied ice problem is the following. Given any 4-regular graph G count the number of orientations ω of G which have the property that at each vertex there are exactly two inward and two outward pointing edges.

Another way of looking at this enumeration problem is as follows. Fix an orientation ω_0 of G. To each directed edge of G assign either a $+1$ or a -1 in such a way that the net flow into each vertex of G is zero. In other words, the ice problem on G is exactly the problem of counting nowhere-zero flows mod 3 in G, discussed in chapters 4 (Appendix) and 9. But this is exactly the evaluation of the Tutte polynomial of G at the point $(0, -2)$, or the rank polynomial of G at $(-1, -3)$.

Equivalently, by using the fact that when G is a planar graph, and G^* is its planar dual, $T(G; x, y)$ equals $T(G^*; y, x)$ and from the relation between Tutte polynomials and chromatic polynomials we see the following.

Proposition 4.6. *The ice problem on a 4-regular planar graph G is equivalent to counting 3-colourings of the dual graph.*

A remarkable result of Lieb (1967) is that if $Z(m, n)$ denotes the ice partition function (that is the number of ice orientations) on the $m \times n$ portion of the square lattice, then

$$\lim_{m,n\to\infty} [Z(m, n)]^{1/mn} = (\tfrac{4}{3})^{3/2} . \tag{4.7}$$

Percus (1971) gives a very complete and clear account of the different approaches to the ice problem culminating in a proof of (4.7) by the transfer matrix method to be described in section 5.

As far as statistical physics is concerned, the problems of most interest are when G is a 3-dimensional lattice. As far as mathematical solution is concerned, only a few 2-dimensional ice models have been solved, a comprehensive account of these is given in Baxter (1982).

The monomer–dimer problem

Let $p(G,k)$ denote the number of k-matchings in the graph G, with the understanding that $p(G,0) = 1$. Define the polynomial $Q(G,z)$ by

$$Q(G,z) = \sum_{k \geq 0} p(G,k) z^{n-2k} .$$

This is a modified form of the matchings polynomial, which is discussed in section 5 of chapter 31 (Godsil). It can also be viewed as the partition function of a physical system.

Consider a collection of sites on the surfaces of a metallic crystal. The surface is exposed to a gas consisting of monomers and dimers, e.g., hydrogen at a high temperature. Each site on the surface is occupied, either by a monomer or by one of the two ends of a dimer. Of course a pair of sites can be occupied by a dimer only if they are neighbours. The state of the system can be represented by a matching in a graph G. This has the crystal sites as its vertices, with two sites adjacent if and only if they can be occupied by the same dimer. Those pairs of sites occupied by dimers determine a matching. Hence the system is completely described by the graph G, the matching and the temperature. (The latter determines the energy gained by filling the crystal sites with monomers and dimers.)

The physical question is whether this system will undergo a phase transition as the temperature varies. In fact it does not, except possibly when there are no monomers. This was proved by Heilmann and Lieb (1972). They showed that all zeros of $Q(G,z)$ have zero real part, and their absolute value is bounded above by $2\sqrt{\Delta - 1}$, where Δ is the maximum valency of a vertex in G. From these facts they eventually deduce the absence of a phase transition. The matchings polynomial has a number of interesting combinatorial properties; see chapter 3 (Pulleyblank).

Hard hexagons

We work on the triangular lattice. Consider a system where some of the vertices of this lattice are covered by hexagons, with each hexagon covering a central vertex and its six neighbours. Two adjacent vertices cannot be both at the centre of a hexagon. We can describe the state σ of the system by assigning a weight 1 to each vertex at the centre of a hexagon, and 0 to the remaining vertices. Thus we may view σ as a 01-vector indexed by the vertices of the lattice. The partition function is

$$Z = \sum_{\sigma} z^{\sigma_1 + \cdots + \sigma_N} \prod_{ij \in E} (1 - \sigma_i \sigma_j) ,$$

where the product is over all edges of the lattice, and the exponent of z is just the number of hexagons. Baxter establishes an invariance property of this partition function using the star–triangle relation. From this he then deduces the free energy. One surprise is an intimate connection with the Rogers–Ramanujan identities. We direct the reader to Andrews (1982) and Andrews et al. (1984) for more information about this relationship.

5. Transfer matrices

Many of the combinatorial problems arIsing in statistical physics can be reduced to enumeration problems, and these in turn can sometimes be solved by the method of *transfer matrices*, which we now discuss.

We begin with the problem of determining the number of ways an $m \times n$ chessboard can be covered with dominoes. Suppose that our board has been covered with dominoes. The given covering can be encoded by assigning one of four states $\{U, D, L, R\}$ to each square of the board. The state of a square determines where the other half of the domino covering the square lies. Thus, if the other half of the domino covering a square is above it, then the square has state U. If it is below we use D, and if it is to the left or the right we use L or R respectively. It should be clear that many assignments of states to squares do not correspond to coverings, but every covering gives rise to a unique assignment of states to squares.

Now we take our coding a step further. View our chessboard as a sequence of n columns. Once a covering is given, the state of the vertices in a given column can be represented by a vector, with states as entries. (Of course, the set of possible vectors for the first and last columns will be a subset of the possible vectors for an interior column.) Thus our covering can now be encoded as a sequence

$$\sigma_1, \ldots, \sigma_n ,$$

where σ_i is the state vector for the ith column.

Let Σ be the set of all possible state vectors for a column. Construct a graph $G = G(\Sigma)$ with vertex set Σ, and with two vertices σ and σ' adjacent if there is a covering such that there are consecutive columns with states σ and σ'. The states that can be taken by the first column form a subset, S say, of $V(G)$ and the states available to the last column form a subset F, say. The number of possible coverings of our $m \times n$ chessboard can now be shown to equal the number of walks of length n in G which start at a vertex in S and finish at a vertex in F. (Our terminology here follows that used in section 5 of chapter 31 by Godsil).

If A is the adjacency matrix of G and we denote the characteristic vectors of the sets S and F by $\chi(S)$ and $\chi(F)$ respectively then the required number of walks is

$$\chi(S)^{\mathrm{T}}A^{n}\chi(F).$$

Using the theory of the spectral decomposition of a symmetric matrix we may write

$$A^{n}=\sum_{\theta}\theta^{n}Z_{\theta},$$

where θ ranges over the distinct eigenvalues of A and Z_{θ} is the matrix representing orthogonal projection onto the eigenspace associated to θ. Denote the largest eigenvalue of A by θ_1. Then by the Perron–Frobenius theory we know that if G is connected then θ_1 is simple, and for any other eigenvalue θ, we have $|\theta|<\theta_1$. It follows that

$$\frac{\chi(S)^{\mathrm{T}}A^{n}\chi(F)}{\theta_1^{n}}\rightarrow\chi(S)^{\mathrm{T}}Z_{\theta_1}\chi(F)$$

and hence that

$$(\chi(S)^{\mathrm{T}}A^{n}\chi(F))^{1/n}\rightarrow\theta_1 \tag{5.1}$$

as n tends to infinity.

The number of domino coverings of our chessboard can be expressed as the number of perfect matchings in a graph, H say. The vertices of H are the squares of the chessboard, and two squares are adjacent in H if and only if they are adjacent on the board. Any domino covering gives a perfect matching in the graph. A generalisation of the original problem can now be obtained as follows. Assign a weight to each edge of H and define the weight of a matching to be the product of the weights of the edges it uses. Instead of simply computing the number of perfect matchings in H, we may determine the sum of the weights of all perfect matchings. The weights we use may be variables, in which case the sum will be a polynomial. (For example we might assign a weight α to each edge of H joining two squares in the same column of the board, and a weight β to the remaining edges. The sum will then be a homogeneous polynomial in α and β.)

In particular, the partition function for the Ising problem itself can be expressed in terms of the number of perfect matchings in an edge-weighted planar graph. (See Appendix E in Thompson 1972.) If we then seek to determine this partition function by a transfer matrix argument, we will find it expressed in the form

$$Z=u^{\mathrm{T}}A^{n}v$$

for a suitable matrix A and vectors u and v. A statistical physicist would then be concerned with properties of the limit

$$(u^{\mathrm{T}}A^{n}v)^{1/n}$$

as n tends to infinity. From the discussion above of the domino problem, we may see that this quantity may be expressed as the largest eigenvalue of a symmetric matrix. Alternatively, we could use the generating function

$$\sum_{n\geq 0} \lambda^n u^{\mathrm{T}} A^n v \,;$$

the largest eigenvalue of A is, in general, the reciprocal of the radius of convergence of this power series.

To close this section, we remark that a solution to the chessboard problem will be found in section 4 of Lovász (1979). (However, it uses Pfaffians rather than transfer matrices. Pfaffians are discussed briefly in section 5 of chapter 31 by Godsil.) A more leisurely introduction to the method of transfer matrices may be found in Percus (1971) and Stanley (1986). A number of applications of this method can be found in Baxter (1982).

6. Duality, stars and triangles

In this section we shall illustrate by example a technique which has been frequently used to resolve (combinatorial) problems of physics. The partition function for the Ising model has two interesting invariance properties. First, if G is a plane graph with dual G^* then their rank polynomials are related by

$$R(G; x, y) = R(G^*; y, x)\,.$$

A proof of this will be found in chapter 9 by Welsh. If

$$\lambda = \left(\frac{\mu^2 + 1}{\mu^2 - 1}\right)^{1/2}$$

then

$$R\left(G; \frac{2}{\lambda^2 - 1}, \lambda^2 - 1\right) = R\left(G; \mu^2 - 1, \frac{2}{\mu^2 - 1}\right) = R\left(G^*; \frac{2}{\mu^2 - 1}, \mu^2 - 1\right).$$

Recalling the relation of the rank polynomial to the partition function described in section 2, this leads to a relation between the partition function for the Ising model on the graph G, expressed in terms of λ, and the partition function of G^*, expressed in terms of $[(\mu^2 + 1)/(\mu^2 - 1)]^{1/2}$. If G is the infinite square lattice then G^* is isomorphic to G and the duality relation becomes an invariance condition. Physic'' n be viewed as a relation between the values of the partition temperatures and low temperatures. For more details, see or Baxter (1982).

ıriance property, we need to consider a generalisation of the ırtition function

$$\mathrm{p}\left[-\sum_{ij\in E(G)} \beta_{ij}\sigma_i\sigma_j\right].$$

Recall from section 2 that in the case of uniform interactions $\beta = J/kT$, here we allow all the interactions J to vary to that β also varies.

By way of example, if G is the square lattice we might have $\beta_{ij} = K$ for all horizontal edges and $\beta_{ij} = L$ for all vertical edges. If $e = ij$ and $\lambda_e = \exp \beta_{ij}$ then we find in place of (2.6) that

$$Z(G) = (\lambda_e - \lambda_e^{-1})Z(G/e) + \lambda_e^{-1}Z(G\backslash e) .$$

Consider the star S and triangle T, with weights as indicated in fig. 6.1.

Suppose that in both cases the vertices 1, 2 and 3 are assigned states σ_1, σ_2 and σ_3. Suppose that the vertices 1, 2 and 3 in S are part of some larger graph G. Then $Z(G)$ is a sum over all possible state assignments of its vertices. For a given state, the corresponding term in the sum is the product of a contribution from the edges of S, and one from the edges not in S. The contribution from S is

$$\exp[\sigma_0(L_1\sigma_1 + L_2\sigma_2 + L_3\sigma_3)] .$$

We divide the possible states into pairs, where members of the same pair differ only in the value of σ_0. Thus we may write $Z(G)$ in the form

$$\sum_{\sigma} 2\cosh(L_1\sigma_1 + L_2\sigma_2 + L_3\sigma_3)f(\sigma) , \tag{6.1}$$

where $f(\sigma)$ is the contribution of the edges not in S, given the values of σ on the vertices of $G\backslash 0$.

Now suppose that we alter G by deleting the vertex 0 and the three edges of S, replacing them with the three edges of T. The new graph, which we denote by G', thus has one less vertex than G. Its partition function can be written in the form

$$\sum_{\sigma} \exp(K_1\sigma_2\sigma_3 + K_2\sigma_1\sigma_3 + K_3\sigma_1\sigma_2)f(\sigma) . \tag{6.2}$$

Then the surprise is, that given L_1, L_2 and L_3, it is possible to choose K_1, K_2 and K_3 so that

$$2\cosh(L_1\sigma_1 + L_2\sigma_2 + L_3\sigma_3) = R\exp(K_1\sigma_2\sigma_3 + K_2\sigma_1\sigma_3 + K_3\sigma_1\sigma_2)$$

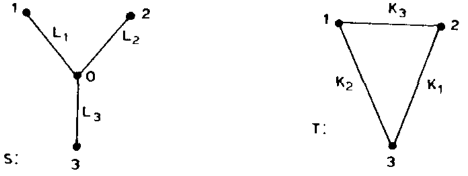


Figure 6.1.

and then $Z(G) = RZ(G')$. To achieve this we need

$$2\cosh(L_1 + L_2 + L_3) = R\exp(K_1 + K_2 + K_3)\,,$$
$$2\cosh(-L_1 + L_2 + L_3) = R\exp(K_1 - K_2 - K_3)\,,$$
$$2\cosh(L_1 - L_2 + L_3) = R\exp(-K_1 + K_2 - K_3)\,,$$
$$2\cosh(L_1 + L_2 - L_3) = R\exp(-K_1 - K_2 + K_3)\,.$$

Denote the four terms on the left by c, c_1, c_2 and c_3 respectively. Then multiplying these four equations together yields that

$$R^4 = cc_1c_2c_3$$

is a necessary condition. Further manipulations yield

$$\sinh(2K_1)\sinh(2L_1) = \sinh(2K_2)\sinh(2L_2) = \sinh(2K_3)\sinh(2L_3) = d^{-1}\,,$$

where

$$d^{-1} = \frac{\sinh(2L_1)\sinh(2L_2)\sinh(2L_3)}{2(cc_1c_2c_3)^{1/2}}\,,$$

as a second necessary condition. If the values of R and K_i are as given by the last three equations then $Z(G) = RZ(G')$. (For help with the missing details, see chapter 6 of Baxter 1982.)

If G is the hexagonal lattice then it can be transformed into a triangular lattice by repeatedly replacing stars by triangles. (The hexagonal lattice is bipartite; replace all the stars centered on vertices in one of the two colour classes.) This gives us a relation between an Ising model on the hexagonal lattice and one on the triangular lattice. We obtain a second, independent, relation by recognIsing that the hexagonal lattice is the planar dual of the triangular lattice. Composing these relations yields an expression for the partition function of an Ising model at low temperature in terms of a partition function at high temperature, for both the triangular and hexagonal lattices. This is an analogue of the relation obtained for the square lattice above. (Again, see Baxter 1982 for more detail.)

Other applications of this star–triangle transformation, which is really a special instance of planar duality theory have been in percolation theory, to the Potts model and to the six and eight vertex ice model. More details may be found in Temperley (1981).

7. Ground states of spin glasses

We will now turn to a different application of combinatorics in statistical physics and outline how some questions about spin glasses can be answered by employing the mathematical machinery of combinatorial optimization. We concentrate on showing that the problem of determining ground states of spin glasses can be viewed as a so-called maximum cut problem in graphs.

The study of order–disorder phenomena is a flourishing branch in today's physics. One of the most successful attempts to understand disorderly systems has been the study of spin glasses. They occupy a central position in this area. The composition of a spin glass is unremarkable – perhaps a few iron atoms scattered in a lattice of copper atoms – but its magnetic properties are confoundedly complicated and sometimes tantalizingly unpredictable. "Spin" is the quantum-mechanical spin from which magnetic effects arise; "glass" refers to the disorder in the orientations and interactions of the spins. For an introduction to the general theory of spin glass models see Mezard et al. (1987).

Physicists have developed a number of theories to model spin glasses and explain their behaviour. Some of these theories predict contradicting phenomena. These phenomena occur in situations which are hard to realize experimentally. In order to test the theories and guide the design of experiments, researchers have designed computer models to simulate the behaviour of spin glasses and then observe which phenomena occur. Some aspects studied in these models lead to optimization problems. The papers by Toulouse (1977), Bieche et al. (1980), and Barahona et al. (1982) have pioneered the study of spin glasses from an optimization point of view and pointed out the close links of the ground state problem to interesting models in combinatorial optimization.

A spin glass is an alloy of magnetic impurities diluted in a non-magnetic material. Alloys that show spin glass behaviour are, for instance, CuMn; the metallic crystal AuFe; the insulator EuSrS; the amorphous metal GdAl. One characteristic of spin glasses is a peak in magnetic susceptibility at a certain temperature. This peak indicates a phase transition. Another phase transition may take place at very low temperature. However, it is an open question at present whether, or under what conditions on the spin glass, such a phase transition occurs, and what order phenomena show up at low temperature.

We will now present a mathematical model of spin glasses. We assume a given spin glass that contains n magnetic impurities (atoms). Each magnetic atom i has a magnetic orientation (spin) which is represented by a 3-dimensional unit-length vector S_i. Between each pair i, j of magnetic atoms there is an interaction J_{ij} that depends on the non-magnetic material and on the distance r_{ij} between the atoms. Several proposals in the literature model this interaction. One common feature of many of these models is that the absolute value of the interaction decreases rapidly with distance and that small changes of distance may result in a change of the sign of the interaction. One example of such an interaction function which is used frequently is

$$J_{ij} = J(r_{ij}) = A\frac{\cos(Dr_{ij})}{B^3 r_{ij}^3},$$

where A, B, and D are material-dependent constants. In another model some number J is chosen and the interactions have to satisfy

$$J_{ij} \in \{0, +J, -J\}.$$

If atoms i and j have spins S_i and S_j, the energy interaction between i and j is given by

$$H_{ij} = J_{ij} S_i \cdot S_j \,,$$

where $S_i \cdot S_j$ denotes the Euclidean inner product. Given a spin configuration or state, σ, the energy of the whole system is given by the Hamiltonian

$$H(\sigma) = -\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} J_{ij} S_i \cdot S_j - h \sum_{i=1}^{n} S_i \cdot F \,,$$

where a unit length vector $F \in \mathbb{R}^3$ represents the orientation of an exterior magnetic field and h represents the strength of this field.

The study of this Hamiltonian is a major issue in statistical physics. Its difficulty has led to considering various simplifications. One such simplification is to replace the 3-dimensional vectors S_i and the magnetic field F by 1-dimensional vectors σ_i, respectively f, with values $+1$ or -1 (called "*Ising spin*"), meaning magnetic north pole "up" and magnetic north pole "down". Such a representation is called the *Ising model* of spin glasses, see section 2 of this chapter for a general introduction to this model and some of its properties. There are, in fact, substances which show an up/down behaviour and for which the Ising model seems to be the "correct" model and not just a simplification.

Models that consider interactions between all pairs of impurities were introduced by Sherrington and Kirkpatrick and are called *long range models*. A number of models consider only interactions between "close" impurities (so-called nearest neighbour interactions), and set to zero the interactions between impurities that are far apart. These models, introduced by Edwards and Anderson, are called *short range models*. Many physicists consider short range models more realistic (see Young 1984). Moreover, a number of substances show short range interactions only: next-neighbour and second-next-neighbour, say.

It is customary to make further simplifications and to consider the spins regularly distributed, say on a 2- or 3-dimensional grid (that is square or cubic lattice). In a typical short range model of such a given structure, interactions are non-zero only along edges of the grid, so, for instance, in two-space, an impurity interacts only with (at most) four other impurities, its neighbours in the grid graph. Two grid models of this type have been studied intensively: the *Gaussian model*, where the interactions are chosen from a Gaussian distribution, and the *$\pm J$-model*, where the interactions between impurities take only the values $+J$ or $-J$, J a fixed positive number, according to some probability distribution. In a real spin glass (an alloy), the magnetic impurities are randomly distributed. Note that in the models just introduced, the spins are regularly distributed in a grid, but the interaction values are considered random.

The partition function of the Ising model has been introduced in section 2. For our purposes, it is useful to write it in the following way. Let Ω be the set of all possible configurations of Ising spins on a grid, say. So $|\Omega| = 2^n$, if there are n spins. Then the behaviour of the system at temperature T is believed to be described by the magnetic partition function

$$Z(T) = \sum_{\sigma \in \Omega} \exp\left(\frac{-H(\sigma)}{kT}\right),$$

where k is the Boltzmann constant. As mentioned in section 2, analytic expressions of the partition function, in general, are not known.

At 0°K, the spin glass system attains a minimum energy configuration. Such a configuration is called a *ground state*. A ground state can be found by minimizing the Hamiltonian associated with the system. We will now present the reduction of the problem of finding a minimum energy spin configuration in the Ising model to a max-cut problem in graphs.

Suppose we have magnetic impurities $1, 2, \ldots, n$ and an exterior magnetic field, 0. We set $V = \{0, 1, \ldots, n\}$ and consider V as the vertex set of a graph $G = (V, E)$. For a pair i, j of impurities, G contains an edge ij if the interaction J_{ij} between i and j is non-zero. An edge $0i$ links every impurity i, $1 \leqslant i \leqslant n$, to the magnetic field 0. Let us call G the *interaction graph* of the spin system. An Ising spin $\sigma_i \in \{-1, +1\}$ is associated with each impurity. The Ising spin σ_0 of the exterior magnetic field can be set to $+1$ without loss of generality. Let h be the strength of the magnetic field and set $J_{0i} = h$ for $i = 1, \ldots, n$, then we can write the Hamiltonian of this model as a quadratic function in ± 1-variables in the following way:

$$H(\sigma) = -\sum_{\substack{ij \in E \\ i,j > 0}} J_{ij}\sigma_i\sigma_j - h\sum_{i=1}^{n} \sigma_i = -\sum_{ij \in E} J_{ij}\sigma_i\sigma_j .$$

Each spin configuration σ corresponds to a partition of V into V^+ and V^-, where $V^+ = \{i \in V \mid \sigma_i = +1\}$ and $V^- = \{i \in V \mid \sigma_i = -1\}$. So we can write the energy of the state σ in the form

$$\begin{aligned} H(\sigma) &= -\sum_{ij \in E(V^+)} J_{ij}\sigma_i\sigma_j - \sum_{ij \in E(V^-)} J_{ij}\sigma_i\sigma_j - \sum_{ij \in \delta(V^+)} J_{ij}\sigma_i\sigma_j \\ &= -\sum_{ij \in E(V^+)} J_{ij} - \sum_{ij \in E(V^-)} J_{ij} + \sum_{ij \in \delta(V^+)} J_{ij} . \end{aligned}$$

Recall that, for each subset W of V, $E(W) = \{ij \in E \mid i, j \in W\}$ and $\delta(W) = \{ij \in E \mid i \in W, j \in V \backslash W\}$ and that the edge sets of type $\delta(W)$ are called cuts. Setting $C = \sum_{ij \in E} J_{ij}$, we see that

$$H(\sigma) + C = 2 \sum_{ij \in \delta(V^+)} J_{ij} ,$$

and defining $c_{ij} = -J_{ij}$ for all $ij \in E$, we find that the problem of minimizing H is equivalent to maximizing

$$c(\delta(V^+)) = \sum_{ij \in \delta(V^+)} c_{ij}$$

over all $V^+ \subseteq V$. The problem of finding, given a graph with edge weights, a cut $\delta(W)$ such that the sum of weights of the edges of $\delta(W)$ is as large as possible is known as the *max-cut problem*. Thus finding a ground state in the Ising model of a spin glass is equivalent to finding an optimum solution of the corresponding max-cut problem.

To determine ground states of spin glasses or, equivalently, cuts of maximum weight, physicists have introduced the so-called *simulated annealing method*. This is an algorithmic analogue of standard techniques in the material sciences where, for instance, large (and perfect) crystals are grown by using a careful scheme of cooling and heating the material to temperatures very close below and above the critical temperature where the liquid freezes into an ordered array of atoms, the crystal. This method was formulated as a general heuristic for the solution of arbitrary combinatorial optimization problems, see, for example, Kirkpatrick et al. (1983), and usually turned out to be a reasonable, though slow, approximation algorithm, see Johnson et al. (1989, 1991).

It soon became clear that, by simulated annealing, states of low energy can be reached but that there is no guarantee or proof that a true ground state can be found. Thus more sophisticated combinatorial methods came into play that we briefly want to mention. More detailed and thorough surveys of these aspects with large lists of references are Barahona et al. (1988), Grötschel et al. (1987).

From the complexity point of view (cf. chapter 29 by Shmoys and Tardos) it turned out that the max-cut problem is NP-hard for general graphs, and so the spin glass problem is. But much more restricted spin systems turned out to lead to hard optimization models. For instance, finding a ground state is NP-hard even if the interaction graph of the spin system is a 3-dimensional grid, or a 3-dimensional grid with just two layers, or even a planar grid with an external magnetic field, provided the interactions are taken from $\{-J, 0, J\}$.

On the other hand, if the interaction graph of the spin system is a planar graph and there is no external field then using the duality theory of planar graphs one can transform the associated max-cut problem to a so-called *Chinese postman problem*. This problem can be solved by the algorithm of Edmonds and Johnson (1973) which combines a series of shortest path calculations and an application of the matching algorithm in an ingenious way. Barahona (1983) extended this to graphs not contractible to the complete graph K_5. Using the Edmonds–Johnson algorithm, ground states of large planar spin systems can (and have been frequently) calculated easily.

The most interesting open questions about spin glasses occur, however, in three dimensions or when an external magnetic field is involved. To solve such (NP-hard) instances various enumeration techniques (e.g., the transfer matrix method) have been designed. The most successful approach seems to be the use of cutting plane algorithms (cf. chapter 37) that are based on an intensive study of the so-called *cut polytope*. This approach is called polyhedral combinatorics and is explained in chapter 30. With these linear programming based cutting plane algorithms, spin systems in three dimensions or two dimensions with magnetic field can be treated that have well over thousand spins. Algorithms of this type terminate with an optimality guarantee, that is, true ground states can be found; and they have further desirable features. But, of course, no polynomial running time guarantee can be given. For more information, see Barahona et al. (1988) and Grötschel et al. (1987), in particular, for a list of open problems in physics that may reach a better level of understanding by a systematic and intensive use of the combinatorial methods outlined above. A collection of papers and surveys on various aspects of spin glasses (including the ones discussed here) is Van Hemmen and Morgenstern (1987).

8. Conclusion

This has been just a glimpse of a fascinating but very difficult area of research. For example there is no mention of the important topic of cluster expansions. Details of this and many other links between combinatorics and physics may be found in the articles of Kasteleyn (1967) and Temperley (1979a). Almost all the problems discussed are probably hard in the sense of computational complexity, except for the very restricted cases. The role of planarity appears to be significant in making a problem easier, see for example Kasteleyn (1961). For more on the complexity of these physical problems see the monograph of Welsh (1993).

As far as solution is concerned, apart from the approach suggested in section 7, the most significant theoretical advance would appear to be the results of Jerrum and Sinclair (1993) that the monomer–dimer problem and the ferromagnetic version of the Ising model have a *fully polynomial randomised approximation scheme*. Put more loosely, this says that there are fast (in the sense of polynomial time) good Monte Carlo methods for these problems. Whether such schemes exist for the other problems discussed in the chapter is an important but as yet unanswered question.

Acknowledgement

We wish to acknowledge some very helpful comments from Laci Lovász and Domokos Szász.

References

Andrews, G.E.
[1982] *q-Series: Their Development and Application in Analysis, Number Theory, Combinatorics, Physics, and Computer Algebra* (American Mathematical Society, Providence, RI).

Andrews, G.E., J. Baxter and P.J. Forrester
[1984] Eight-vertex SOS model and generalized Rogers–Ramanujan identities, *J. Statist. Phys.* **35**, 193–266.

Barahona, F.
[1983] The max cut problem in graphs not contractible to K_5, *Oper. Res. Lett.* **2**, 107–111.

Barahona, F., R. Maynard, R. Rammal and J.P. Uhry
[1982] Morphology of ground states of a two dimensional frustration model, *J. Phys. A* **15**, L673-L699.

Barahona, F., M. Grötschel, M. Jünger and G. Reinelt

[1988] An application of combinatorial optimization to statistical physics and circuit layout design, *Oper. Res.* **26**, 493–513.

Baxter, R.J.

[1982] *Exactly Solved Models in Statistical Mechanics* (Academic Press, London).

Bieche, I., R. Maynard, R. Rammal and J.P. Uhry

[1980] On the ground states of the frustration model of a spin glass by a matching method of graph theory, *J. Phys. A* **13**, 2553–2576.

Biggs, N.

[1977] *Interaction Models* (Cambridge University Press, Cambridge).

Broadbent, S.R., and J.M. Hammersley

[1957] Percolation processes I. Crystals and mazes, *Proc. Cambridge Philos. Soc.* **53**, 629–641.

Cipra, B.A.

[1987] An introduction to the Ising model, *Amer. Math. Monthly* **94**, 937–959.

Edmonds, J., and E.L. Johnson

[1973] Matching, Euler tours and the Chinese postman, *Math. Programming* **5**, 88–124.

Fisher, M.E.

[1961] Critical probabilities for cluster size and percolation problems, *J. Math. Phys.* **2**, 620–627.

Fortuin, C.M., and P.W. Kasteleyn

[1972] On the random-cluster model. I. Introduction and relation to other models, *Physica* **57**, 536–564.

Grimmett, G.R.

[1989] *Percolation* (Springer, New York).

Grötschel, M., M. Jünger and G. Reinelt

[1987] Calculating exact ground states of spin glasses, a polyhedral approach, in: *Heidelberg Colloquium on Spin Glasses*, eds. J.L. van Hemmen and I. Morgenstern (Springer, Berlin) pp. 325–353.

Hammersley, J.M.

[1961a] Comparison of atom and bond percolation processes, *J. Math. Phys.* **2**, 728–733.

[1961b] The number of polygons on a lattice, *Proc. Cambridge Philos. Soc.* **57**, 516–523.

Hammersley, J.M., and D.J.A. Welsh

[1965] *First Passage Percolation, Subadditive Processes, Stochastic Networks and Generalised Renewal Theory, Bernoulli-Bayes-Laplace Anniversary Volume* (Springer, Berlin) pp. 61–110.

Hara, T., and G. Slade

[1992] The face expansion for self avoiding walk in five or more dimensions, *Rev. Math. Phys.* **4**, 235–327.

Harris, T.E.

[1960] A lower bound for the critical probability in a certain percolation process, *Proc. Cambridge Philos. Soc.* **56**, 13–20.

Heilmann, C.J., and E.H. Lieb

[1972] Theory of monomer-dimer systems, *Comm. Math. Phys.* **28**, 190–232.

Jerrum, M.R., and A.J. Sinclair

[1993] Polynomial-time approximation algorithms for the Ising model, *SIAM J. Comput.* **22**, 1087–1116.

Johnson, D.S., C.R. Aragon, L.A. McGeoch and C. Schevon

[1989] Optimisation by simulated annealing: an experimental evaluation. I. Graph partitioning, *Oper. Res.* **37**, 865–892.

[1991] Optimisation by simulated annealing: an experimental evaluation. II. Graph colouring and number partitioning, *Oper. Res.* **39**, 378–406.

Kasteleyn, P.W.

[1961] The statistics of dimers on a lattice, *Physica* **27**, 1209–1225.

[1967] Graph theory and crystal physics, in: *Graph Theory and Theoretical Physics*, ed. F. Harary (Academic Press, London) pp. 43–110.

Kesten, H.

[1963] On the number of self avoiding walks, *J. Math. Phys.* **4**, 960–969.

[1980] The critical probability of bond percolation on the square lattice equals $\frac{1}{2}$, *Comm. Math. Phys.* **74**, 41–59.
[1982] *Percolation Theory for Mathematicians* (Birkhäuser, Boston).

Kingman, J.F.C.
[1973] Subadditive ergodic theory, *Ann. Probab.* **1**, 833–909.

Kirkpatrick, S., C.D. Gelatt and M.P. Vecchi
[1983] Optimization by simulated annealing, *Science* **220**, 671–680.

Lieb, E.H.
[1967] Residual entropy of square ice, *Phys. Rev.* **162**, 162–171.

Lovász, L.
[1979] *Combinatorial Problems and Exercises* (North-Holland, Amsterdam). 2nd Edition: 1993.

Madras, N., and G. Slade
[1993] *The Self Avoiding Walk* (Birkhäuser, Boston).

Mezard, M., G. Parisi and M.A. Virasoro
[1987] *Spin Glass Theory and Beyond* (World Scientific, Singapore).

Oxley, J.G., and D.J.A. Welsh
[1979] The Tutte polynomial and percolation, in: *Graph Theory and Related Topics*, eds. J.A. Bondy and U.S.R. Murty (Academic Press, London) pp. 329–339.

Percus, J.K.
[1971] *Combinatorial Methods* (Springer, New York).

Ruelle, D.
[1969] *Statistical Mechanics* (Benjamin, New York).

Smythe, R.T., and J.C. Wierman
[1978] First passage percolation on the square lattice, *Lecture Notes in Mathematics*, Vol. 671 (Springer, Berlin).

Stanley, R.P.
[1986] *Enumerative Combinatorics*, Volume I (Wadsworth and Brooks/Cole, Monterey, CA).

Sykes, M.F., and J.W. Essam
[1964] Exact critical percolation for size and bond problems in two dimensions, *J. Math. Phys.* **5**, 1117–1127.

Temperley, H.N.V.
[1979a] Graph theory and continuum statistical mechanics, in: *Applications of Graph Theory*, eds. R.J. Wilson and L.W. Beineke (Academic Press, London) pp. 121–148.
[1979b] Lattice models in discrete statistical mechanics, in: *Applications of Graph Theory*, eds. R.J. Wilson and L.W. Beineke (Academic Press, London) pp. 149–175.
[1981] *Graph Theory and Applications* (Ellis Horwood, Chichester).

Thompson, C.J.
[1972] *Mathematical Statistical Mechanics* (MacMillan, New York).

Toth, B.
[1985] A lower bound on the critical probability of the square lattice site percolation, *Z. Wahrscheinlichkeitstheorie u. Verw. Gebiete* **69**, 19–22.

Toulouse, G.
[1977] Theory of the frustration effect in spin glasses: I, *Comm. Phys.* **2**, 115–119.

van Hemmen, J.L., and I. Morgenstern
[1987] *Heidelberg Colloquium on Spin Glasses* (Springer, Berlin).

Welsh, D.J.A.
[1990] The computational complexity of problems from statistical physics, in: *Disorder in Physical Systems*, eds. G.R. Grimmett and D.J.A. Welsh (Oxford University Press, Oxford) pp. 305–321.
[1993] *Complexity: Knots, Colourings and Counting, London Mathematical Society Lecture Note Series*, Vol. 186 (Cambridge University Press, Cambridge).

Whittington, S.G., and C.E. Soteros
[1990] Lattice animals: rigorous results and wild guesses, in: *Disorder in Physical Systems*, eds. G.R. Grimmett and D.J.A. Welsh (Oxford University Press, Oxford) pp. 321–334.

Wierman, J.C.
[1981] Bond percolation on honeycomb and triangular lattices, *Adv. in Appl. Probab.* **13**, 293–313.
Young, A.P.
[1984] Spin glasses, *J. Statist. Phys.* **34**, 871–881.
Zuev, A.S.
[1987] Estimates for the percolation threshold for a square lattice (in Russian), *Theor. Probab. Appl.* **32**, 551–552.

CHAPTER 38

Combinatorics in Chemistry

Dennis H. ROUVRAY

Department of Chemistry, The University of Georgia, Athens, GA 30602, USA

Contents

HANDBOOK OF COMBINATORICS
Edited by R. Graham, M. Grötschel and L. Lovász

1. General introduction

The science of chemistry concerns itself basically with the study of molecules, imperceptibly small architectural edifices of atoms held together by chemical bonds. There are several different types of bond ranging from the very strong to the very weak. Much of the work of chemists is involved with the making or breaking of chemical bonds in molecular structures. Whenever such bonds are formed or eliminated, the molecule in question is said to undergo a chemical transformation. To an ever increasing extent, chemists are finding it convenient to represent both molecules themselves and the various transformations they undergo by means of chemical graphs. In the case of molecules, the graphs are commonly referred to as *molecular graphs*; those used to represent chemical transformations are called *reaction graphs*. Because all chemical systems, including individual molecules, possess both metric (geometric) and nonmetric (topological) attributes, graphs alone cannot reflect the totality of their chemical behavior. However, graphs are especially useful for characterizing the nonmetric attributes of molecules, which are today generally recognized as being of at least equal importance to the metric ones.

The exploitation of graphs in a chemical context brings in its wake a number of interesting consequences, some of which are worthy of brief consideration here.

(i) Certain of the results obtained by graph theorists and combinatorialists are directly applicable to the solution of chemical problems. An example of this is provided by the extensive use of Pólya's enumeration theorem for the enumeration of many different types of chemical isomers (see section 3).

(ii) Specific problems encountered by chemists who make use of graphs may present new challenges to the mathematical community. As an example, we cite the need to examine and classify the more than 150 different scalar numerical graph invariants that chemists are currently employing to characterize chemical species.

(iii) There exists the possibility that solutions to chemical problems found by chemists may uncover new mathematical knowledge. Here our example is provided by the proof of the so-called *pairing theorem* in a chemical context (see section 5).

(iv) The increasing use of graph-theoretical and combinatorial methods in chemistry means that, as more chemists become aware of the power and elegance of these methods, further research into chemical applications will be stimulated and make chemistry into an increasingly important proving ground for such methods. We mention in passing that there are already more applications in chemistry than in any other scientific discipline.

(v) To cope with the mounting number of publications in the areas of chemical graph theory and chemical combinatorics, new outlets will probably be needed as the existing journals become overloaded. The pressure of publications has so far spawned two new interdisciplinary journals to keep pace with papers being generated in this area. The journals are *Mathematical Chemistry* (known affectionately as *Match*) founded in 1975, and the *Journal of Mathematical Chemistry* founded in 1987.

A key question pertaining to the use of graphs in the chemical domain is whether they contain sufficient information to function as effective descriptors of chemical systems. Most of the graphs used by chemists to date have been planar, loopless and nondirected. The question ultimately resolves itself into whether the graph connectivity (usually referred to by chemists as the graph *topology*) of such graphs yields enough insights to make possible prediction of the behavior of systems of chemical interest. This issue has been a recurring and controversial one. Workers in chemical graph theory and chemical combinatorics have consistently maintained that graphs are highly relevant in chemistry (Trinajstić 1983). Others have contended that graphs have a role to play only in fringe chemical problems. What cannot be doubted, however, is the remarkable upsurge in the number of publications in the field in recent years. The decades of the 1970s and 1980s witnessed a steady increase of around 25% per annum in the output of papers, and this trend seems to be continuing into the 1990s. Among this wealth of material are to be found comprehensive review articles (Rouvray 1974, Balaban 1976, Rouvray and Balaban 1979, Balasubramanian 1985, Balaban 1985), conference proceedings (King 1983, King and Rouvray 1987, Klein and Randić 1990), and specialist monographs (Balaban 1976, Trinajstić 1983, Kennedy and Quintas 1988, Merrifield and Simmons 1989, Johnson and Maggiora 1990, Rouvray 1990b, Bonchev and Rouvray 1991, 1992). Chemical graph theory is clearly a burgeoning field of scientific research activity at the present time.

2. Early uses of graphs

To understand how this came about, we trace briefly the origins of the current interest in graphs by the chemical community. Graphs have been used in the physical sciences for over two centuries and some of the earliest uses were in a chemical setting (Rouvray 1990a). For instance, the various interactions between the pairs of molecules in sets of molecules undergoing chemical transformations known as double decompositions were represented by reaction graphs as early as 1768. The molecules were depicted as the nodes and the interactions as the edges of the graph; an example of such a graph is shown in fig. 2.1(a). The first depictions of the structure of individual molecules were made by William Higgins in 1789. Here the nodes now represent the atoms and the edges the supposed forces holding the atoms together, as seen in fig. 2.1(b). One of the earliest representations of the benzene ring (C_6H_6), reproduced in fig. 2.1(c), is to be found among the 368 graphical depictions of chemical structures by Loschmidt in 1861. Another pioneer in the early use of graphical formulas was Crum Brown, who also in 1861 first drew formulas such as that shown in fig. 2.1(d). The use of graphs to depict the structure of molecules has been the most common application of graph theory to date in the chemical domain.

It is interesting to observe here that the word graph itself is of chemical origin.

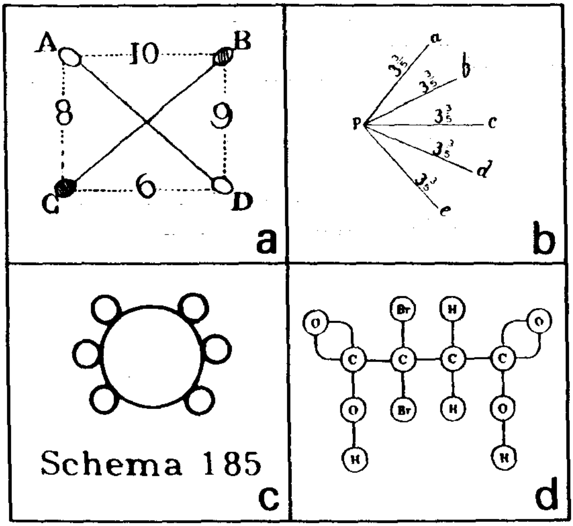


Figure 2.1. Some early graph-theoretical representations of molecules dating from the years (a) 1768, (b) 1789, (c) 1861 and (d) 1861.

The word actually derives from the *graphic notation* of the chemists of the last century – a term that was widely used to denote the structures of molecules such as that shown in fig. 2.1(d) (Crum Brown 1865). The mathematician Sylvester first proposed that the abbreviated form *graph* be employed to describe these chemical structures (Sylvester 1878). From this point on, the word graph appears with increasing frequency in the mathematical literature. It did not catch on immediately, however, among the chemical community. Chemists began to refer to graphic notation as the *structural formula*, and this eventually evolved into the now more current *constitutional formula*. The term *tree* was first introduced into the chemical literature somewhat earlier when the mathematician Cayley published a chemical paper describing a method for the enumeration of various classes of molecule that constitute the members of homologous series (Cayley 1875). A homologous series is a series of related molecules in which each member is represented by a general formula and successive members differ from each other by a methylene (CH_2) unit. More on the early history of graph theory and combinatorics is to be found in chapter 44 by Biggs et al. on "The History of Combinatorics".

Determinations of the permissible structures that molecules may possess rely on two basic combinatorial formulas. The first is the formula giving the number of independent cycles or cyclomatic number of a graph, and is usually expressed as

$$\mu(G) = m - n + 1 , \qquad (2.1)$$

where m and n are, respectively, the numbers of edges and vertices in the graph

G. When employed in a chemical context, $\mu(G)$ includes the multiple bonds present in molecular species. Thus, a double bond (represented by a graph-theoretical 2-cycle) becomes a single cycle and a triple bond two cycles. In fact, the cyclomatic number can be used to define the extent of saturation in hydrocarbon and other molecules (Rouvray 1975). The second combinatorial formula used is the well-known Euler polyhedral formula:

$$n - m + \phi = 2\,, \tag{2.2}$$

where ϕ represents the number of faces in the polyhedron. Normally only convex polyhedra that can be mapped onto the sphere are considered. Both of these formulas have found extensive use in the chemical domain in tasks ranging from the computer generation of isomeric structures (Golender and Rozenblit 1983), through identification of chemically favored coordination compounds (King 1969), to assessment of the inherent rigidity in molecular polyhedral isomerizations (King 1988).

3. Isomer enumeration techniques

Although now primarily of historical interest, graph-theoretical and combinatorial techniques have proven themselves indispensable in the enumeration of chemical isomers. This application represents the earliest use of these techniques for the solution of computational chemical problems. An isomeric pair of molecules consists of two molecules of the same atomic constitution that differ in at least some of their properties. Chemical species can exhibit many different types of isomerism and there exist numerous classification schemes for isomers (Slanina 1986). The two fundamental classes of isomers are the *constitutional isomers* (also known as structural isomers) and *stereoisomers*. Constitutional isomeric pairs differ in their atomic connectivity whereas stereoisomeric pairs differ in the relative positioning of their atoms in space, the different ways of positioning four different substituents around a tetrahedral carbon atom being a typical example. All isomeric species should be stable for periods of time long in comparison to those in which measurements of their properties are made. Isomer enumeration yields not only isomer counts but also affords a means of estimating the number of stationary points on the potential energy hypersurfaces used to characterize molecular structures. This renders the enumeration of chemical graphs a valuable adjunct to theoretical studies on chemical reactivity (Slanina 1986). A more detailed discussion on isomers and their enumeration is to be found in the surveys of Rouvray (1974), Balaban (1985) and Simon (1987), and in the book of Slanina (1986).

Broadly speaking, the now mature field of isomer enumeration may be viewed as having passed through four principal phases, which have involved the successive exploitation of (i) recursion formulas, (ii) generating functions, (iii) the enumeration theorem of Pólya (1937), and (iv) more recent techniques, such as the double coset method of Ruch et al. (1970) and the reformulated method of

Redfield (Redfield 1927, Davidson 1981, Lloyd 1988). The area as a whole illustrates perhaps better than any other in chemical combinatorics the close interplay and mutual stimulation that has existed between the mathematical and chemical communities for well over a century. The net result of this long collaboration in the enumeration of chemical isomers is that at this stage virtually all species of chemical interest have now been enumerated. One notable exception here concerns the polycyclic aromatic hydrocarbons that may be represented as planar hexagonal animals, i.e., planar graphs with no cut vertices in which every interior region is a hexagonal unit cell. No general enumerative procedure for such graphs exists, though computations have been made for polyhexes containing up to sixteen hexagons (Knop et al. 1990). Over the past few years tables of isomer counts for various classes of molecules have begun to appear (Knop et al. 1985, Dias 1987).

The earliest use of a combinatorial technique to determine isomer counts was that of the chemist Flavitsky in 1871. He studied the homologous series of the alcohols and enumerated the first ten members. The general formula for the alcohols is $C_nH_{2n+1}OH$ and the members are subdivided into three classes known as primary, secondary and tertiary alcohols. Class assignment is made on the basis of the number of carbon atoms attached to the carbon (C) atom bearing the hydroxyl (OH) group. For primary alcohols this number is one, for secondary alcohols two, and for tertiary alcohols three; one typical member of each class is illustrated in fig. 3.1. In mathematical terms the problem is equivalent to enumerating rooted tree graphs all of whose vertices are of valence one or four. Flavitsky made use of recursion relations to enumerate the alcohols; examples of

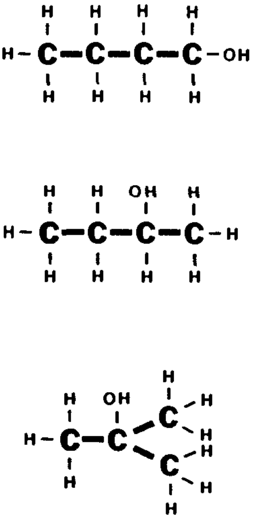


Figure 3.1. Examples of a primary, secondary and tertiary alcohol.

the kinds of relation he employed are:

$$p_n = T_{n-1} \tag{3.1}$$

and

$$\begin{aligned} s_n &= T_1T_{n-2} + T_2T_{n-3} + \cdots + T_{(n-2)/2}T_{n/2} \quad (n \text{ even}), \\ s_n &= T_1T_{n-2} + T_2T_{n-3} + \cdots + T_{(n-3)/2}T_{(n+1)/2} \\ &\quad + \tfrac{1}{2}[T_{(n-1)/2}(1 + T_{(n-1)/2})] \quad (n \text{ odd}), \end{aligned} \tag{3.2}$$

where p_n, s_n and T_n are respectively the number of primary alcohols, the number of secondary alcohols, and the total number of alcohols of all kinds containing n carbon atoms. This type of approach was greatly extended in the 1930s when the chemists Henze and Blair developed recursion formulas for the enumeration of many different homologous series, including the alkanes, alcohols, amines, ethers and organic acids (see Rouvray 1974). Several further elaborations of this work have been made in recent years (see Balaban 1985).

Generating functions were introduced into isomer enumeration studies by the mathematician Cayley. In 1857 he had enumerated rooted trees by means of a generating function of the form:

$$(1-x)^{-1}(1-x^2)^{-A_1}(1-x^3)^{-A_2}\cdots(1-x^n)^{-A_{n-1}} = 1 + A_1x + A_2x^2 + \cdots, \tag{3.3}$$

where x is a variable, n the number of vertices, and A_{n-1} is the coefficient of x^{n-1} which gives the number of rooted trees on n vertices. In 1874 Cayley adapted this function to the enumeration of unrooted trees; such trees are equivalent to members of the alkane homologous series, C_nH_{2n+2}. By 1875 he had published a long paper on enumerating trees of maximal valence four (alkanes), three and two (Cayley 1875). Cayley succeeded in enumerating up to the thirteenth member of the alkane series, though two of his results were later shown to be in error. Modern isomer counts for alkane molecules containing up to 50 carbon atoms and also for alcohol molecules containing up to 40 carbon atoms based on the compilations of Knop et al. (1985) are presented in tables 3.1 and 3.2 respectively.

A major step forward in the enumeration of graphs, including many graphs of chemical interest, was made by Pólya (1937). The key part of this work, originally known as the *Hauptsatz*, is nowadays universally referred to as the enumeration theorem of Pólya or Redfield–Pólya, to commemorate the fact that much of Pólya's work had been adumbrated in Redfield's classic paper (Redfield 1927, Lloyd 1988). Pólya's complete paper was translated into English exactly 50 years after its appearance (Pólya and Read 1987); by then its relevance in the chemical world had been widely recognized. Pólya also published several follow-up papers illustrating how his theorem could be applied to the enumeration of molecules (see Rouvray 1974). The enumeration theorem rests essentially on an integrated use of symmetry classes of graphs, generating functions and weighting factors. In

Table 3.1
Enumeration of members of the alkane homologous series, C_nH_{2n+2}

Value of n	Isomer count
1	1
2	1
3	1
4	2
5	3
6	5
7	9
8	18
9	35
10	75
20	366319
30	4111846763
40	62481801147341
50	1117743651746953270

Table 3.2
Enumeration of members of the alcohol homologous series, $C_nH_{2n+1}OH$

Value of n	Primary	Secondary	Tertiary	Total
1	1	0	0	1
2	1	0	0	1
3	1	1	0	2
4	2	1	1	4
5	4	3	1	8
6	8	6	3	17
7	17	15	7	39
8	39	33	17	89
9	89	82	40	211
10	211	194	102	507
20	2156010	2216862	1249237	5622109
30	35866550869	37977600390	22147214029	95991365288
40	720807976831447	773973501324306	459220572506066	1954002050661819

concise form, the theorem may be expressed by the equation

$$C(x) = |\mathcal{G}|^{-1} \sum_{(j)} g_{(j)} \prod_{k=1}^{s} f^{j_k}(x^k) , \tag{3.4}$$

where $C(x)$ is a configuration counting series, $f(x)$ a figure counting series, $\mathcal{G}$ is a permutation group of degree q and order g, $g_{(j)}$ is the number of elements of $\mathcal{G}$ of type (j) with $(j) = (j_1, j_2, \dots, j_q)$, j_k the number of cycles of length k with $k = 1, 2, \dots, q$, and the summation extends over all partitions (j) of q subject to the condition

$$j_1 + 2j_2 + \cdots + qj_q = q . \tag{3.5}$$

By making use of the cycle index of the permutation group $\mathcal{G}$ arising from some enumeration problem, Pólya's approach reduces enumeration of the possible combinatorial configurations to enumeration of their equivalence classes. The cycle index is defined as follows:

$$Z(G) = |\mathcal{G}|^{-1} \sum_{(j)} N_{(j)} f_1^{j_1} f_2^{j_2} \cdots f_s^{j_s} , \tag{3.6}$$

where $N_{(j)}$ is the number of elements of $\mathcal{G}$ of cycle type (j) and the f terms are indeterminates. As an example of a cycle index of chemical relevance, we consider that for the benzene molecule, C_6H_6, which is comprised of six carbon atoms in the form of a regular hexagon with a hydrogen atom attached to each carbon. Its permutation group $\mathcal{G}$ is the point group D_6 consisting of twelve different elements that generate the cycle index for benzene:

$$Z(C_6H_6) = \tfrac{1}{12}(f_1^6 + 4f_2^3 + 3f_1^2 f_2^2 + 2f_3^2 + 2f_6) . \tag{3.7}$$

Substitution of an appropriate generating function to replace the f indeterminates transforms (3.7) into a generating function for nonequivalent configurations. In the case of progressive replacement of the hydrogen atoms in benzene by a single, monovalent substituent X the appropriate function to substitute is

$$f(x) = 1 + x . \tag{3.8}$$

Substitution of (3.8) into (3.7) yields the configuration counting series for the benzene molecule:

$$C(x) = 1 + x + 3x^2 + 3x^3 + 3x^4 + x^5 + x^6 . \tag{3.9}$$

The coefficients of $C(x)$ give the numbers of substitutional isomers obtainable when the hydrogen atoms in benzene are progressively replaced by X. The complete set of substitutional isomers is illustrated in fig. 3.2.

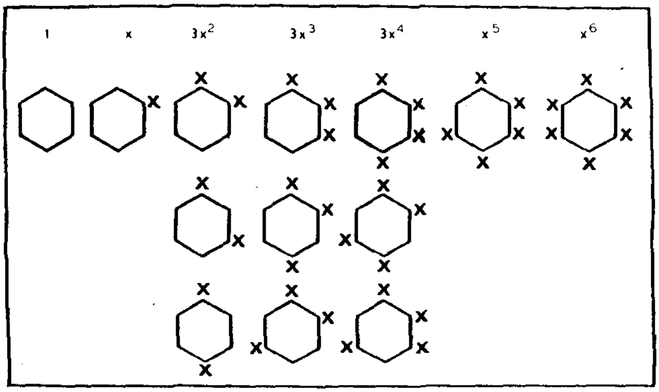


Figure 3.2. The substitution products obtained upon progressively substituting a benzene ring (C_6) with a monovalent substituent.

Pólya's enumeration theorem has been employed to determine isomer counts for numerous systems of chemical interest (Rouvray 1974, Balaban 1985, Balasubramanian 1985). Major contributions to this area have been made by Balaban and his co-workers (Balaban 1985). Solutions have been found for the enumeration among others of stereoisomers, valence isomers, cubic graphs and constitutional isomers. Valence isomers are a subclass of constitutional isomers for which the hydrogen-depleted molecular graphs have identical partitioning of their vertex degrees. A number of the problems encountered in this work relate to the well-known "necklace problem" in combinatorics; comprehensive coverage of this area is to be found in a three-volume monograph by Balaban (1987). The application of Pólya's theorem to the enumeration of spectroscopic signals in nuclear magnetic resonance studies on molecules has been discussed by Balasubramanian (1985). In this case the number of signals is given by the number of magnetic equivalence classes for the molecules of interest. The converse problem to isomer enumeration, namely determination of the extent to which it is possible to deduce molecular symmetry from the spectrum of isomer counts for a given chemical species, has been addressed by Hässelbarth (1987).

Starting in the early 1970s interest began to be focused on the enumeration of so-called nonrigid molecular species, that is molecules which change their conformation or undergo other intramolecular changes during the time scale of experiments carried out on them. Such studies involve separate considerations of (i) the underlying molecular framework, and (ii) the relative positions of the substituents X attached to the framework. Both have given rise to some interesting combinatorial problems (Maruani and Serre 1983, Brocas 1986). In case (i) the symmetry point group appropriate for the idealized rigid framework will not correctly characterize the symmetry of the actual structure. It is necessary to employ a new group to allow for the additional symmetry elements introduced by the various intramolecular motions. In case (ii) the isomers formed as a result of substituent interchanges are referred to as *permutational isomers*. Whenever a new isomer is formed in this way, a *permutational isomerization reaction* is said to have taken place. The total possible number of such reactions for a molecule with n attached identical X substituents will be $n!$, though this number is not attained for molecules possessing a point symmetry. The $n!$ isomerization reactions will be reduced because (i) some of the permutations will not generate new isomers as they represent proper rotations of the molecular point group, and (ii) all the X substituents are identical and distinguishable only by labels, so some of the isomerization reactions will necessarily be indistinguishable.

Several different procedures have been developed to enumerate the isomers arising from nonrigid molecular species. The most general way, however, remains that of Redfield (1927). When expressed in modern terminology, his method is seen as bringing together in one powerful generalization the disparate strands of various enumerative procedures introduced by Pólya and later workers down to the present time (Davidson 1981). Redfield's approach envisaged m sets $\mathscr{X}_1$, $\mathscr{X}_2, \ldots, \mathscr{X}_m$, each of which contained n elements arranged in horizontal n-tuples to form $m \times n$ matrices. Those matrices that are column equivalent are said to have the same correspondence in their elements. In all, there are $(n!)^m$ such matrices and $(n!)^m/n!$ correspondences. The problem is to enumerate the number of injective mappings whereas Ruch, Klemperer and Brown and co-workers (see Davidson 1981) used double cosets. Double cosets have been the focus of much interest in the chemical literature of the past two decades; there is even a review of their various applications in the physical sciences (Ruch and Klein 1983). Double cosets partition groups in terms of pairs of subgroups. If subgroup A of the symmetric group, S_ν, of degree ν contains all the symmetry permutations arising from rotation of a molecular framework as a whole plus the intramolecular rotations of that framework, and subgroup $B \subseteq S_\nu$ contains all the permutations of identical substituents X on the framework attachment sites, then the double coset $A\ g\ B$ will consist of the permutations:

$$A\ g\ B = \{a\ g\ b \mid a \in A, b \in B\}\ , \quad g \in S_\nu\ . \tag{3.10}$$

This particular formalism has been employed not only in the enumeration of permutational isomers but also for polytopal rearrangements and the chemical reactions of molecules (Brocas et al. 1984).

4. The study of reaction networks

As mentioned in our historical introduction, the earliest use of graphs in a chemical context was for depiction of the mutual interactions of sets of molecules. Such usage has become commonplace in recent decades, even though it is now realized that graphs or even hypergraphs are not necessarily the best mathematical constructs to represent chemical reactions. A *chemical reaction* may be defined as sequence of elementary steps $\{\zeta_k\}$ in each of which a set of molecules is chemically transformed into some other set. A linear combination of such steps of the general form

$$M = a_1\zeta_1 + a_2\zeta_2 + \cdots + a_k\zeta_k\ , \tag{4.1}$$

where the a_k are real coefficients, is referred to as a *mechanism* of the reaction. A collection of single-step chemical reactions that are interlinked by the flow of molecules from one reaction to another are said to constitute a *reaction network*. A reacting chemical system can thus be viewed as a linear transformation from a k-dimensional real vector space generated by the elementary steps into an A-dimensional real vector space generated by the molecules (Milner 1964, Sellers 1984).

A reacting system can be adequately represented by a graph only in certain special instances. An example of such an instance occurs when every elementary step in the network is an isomerization reaction. In this case, two molecules, α and β, undergoing reaction will be governed by the chemical equation $\alpha = \beta$. Every species can then be represented by a graphical vertex and every reaction by an edge or sequence of edges of a graph. Since isomerization reactions involve only the molecular framework and the various transformations it undergoes, the substituents can be neglected to a first approximation. Moreover, because the initial and final geometries of the framework remain identical in such reactions, the distinguishability of isomerization reactions will be determined by the condition

$$r_i = h_k \cdot r_j \cdot h_k^{-1}\,, \tag{4.2}$$

where r_i and r_j represent isomerization reactions carried out in a totally symmetric environment, and $h_k \in \mathscr{H}$, the permutation group that acts on the indices of the framework attachment sites, and a subgraph of S_ν. If condition (4.2) holds, the reactions r_i and r_j are indistinguishable, whereas if it does not the reactions are distinguishable. For a molecule in a less symmetric environment, the full point group cannot be used in this way to define reaction distinguishability. In such cases this group is replaced by a subgroup defining the symmetry of the site that the molecule occupies (Klemperer 1972).

The condition (4.2) is described as a *conjugacy relation*. Each of the conjugacy classes generated in S_ν with respect to the group $\mathscr{H}$ will represent a set of indistinguishable permutational isomerization reactions in a totally symmetric environment, provided we neglect those classes comprised of elements of the proper rotational symmetry group $\mathscr{P}$ for molecules in a chiral environment. The number of conjugacy classes in S_ν with respect to $\mathscr{H}$ minus the number of classes comprised of elements of $\mathscr{P}$ yields the number of formally distinguishable reactions in a symmetric environment. Klemperer (1972) expressed this result mathematically by developing a formula for counting the number of conjugacy classes in S_ν with respect to $\mathscr{H}$. A counting polynomial was defined as:

$$C(\mathscr{H}; a_1, a_2, \ldots, a_n) = \sum_{(j_1, j_2, \ldots, j_n)} A_{j_1 j_2 \cdots j_n} a_1^{j_1} a_2^{j_2} \cdots a_n^{j_n}\,, \tag{4.3}$$

where $A_{j_1 j_2 \cdots j_n}$ is the number of conjugacy classes in S_ν with respect to $\mathscr{H}$ of cycle type $(j_1, j_2, \ldots, j_n)$ and the a terms are indeterminates. The number of distinguishable reactions is given by the formula

$$A'_{j_1 j_2 \cdots j_n} = A_{j_1 j_2 \cdots j_n} - N_{j_1 j_2 \cdots j_n}\,, \tag{4.4}$$

where $N_{j_1 j_2 \cdots j_n}$ is the number of conjugacy classes in $\mathscr{H}$ of cycle type $(j_1, j_2, \ldots, j_n)$ and A' pertains to the permutations generated by proper rotations of the molecular framework.

As indicated, reaction networks cannot in general be adequately represented by graphs or 1-dimensional simplicial complexes. In this respect they differ markedly from electrical networks, a fact that might be anticipated in view of the complexity of most chemical reactions. By redefining a reaction network as a finite set of chemicals together with a finite set of reactions such that there exist a homomorphism $\partial\colon \mathscr{C}_1 \to \mathscr{C}_2$, where $\mathscr{C}_1$ and $\mathscr{C}_2$ are respectively the free abelian groups generated by the chemicals and the reactions, Sellers (1967) was able to establish the following result:

$$\gamma_2 - \gamma_1 = \eta_2 - \eta_1 , \tag{4.5}$$

where γ_1 and γ_2 are the respective numbers of chemicals and reactions, η_1 is the maximum number of linearly independent conservation conditions, and η_2 the maximum number of stationary states. The algebraic complexes that can be used to represent reaction networks in which there are no isomerization reactions have been discussed in detail by Sellers (1967).

An interesting combinatorial problem that arises in the study of reaction mechanisms is the construction of the complete listing of the chains of steps by which a specified chemical reaction can proceed in a given reaction network (Sellers 1989). The problem is analogous to, though more difficult than, finding all the paths connecting a pair of vertices in a graph. A definition applicable both to graph paths and reaction mechanisms is that a mechanism is said to be *direct* whenever no distinct mechanism for the reaction can be formed from a subset of its steps. This implies that a mechanism is direct iff no cycle can be found from a subset of its steps. Sellers (1989) has established that any reaction mechanism can be decomposed into a sum of irreducible direct mechanisms. Although such a decomposition will not in general be unique, the set of all direct mechanisms for a specified reaction will be a unique characteristic of a reaction network. Because direct mechanisms are not in general linearly independent, any list of supposedly distinct linear combinations will contain repetitions. Moreover, the list will be infinitely long unless the set of all such mechanisms is separated into equivalence classes.

Sellers (1989) developed a systematic way of determining all the possible direct mechanisms for one or more overall reactions in a reaction network based on concepts adapted from algebraic topology. Each possible type of mechanism was characterized by a convex subset of a finite-dimensional vector space. Use was made of geometric constructions in which mechanisms were modeled by a finite set of polygonal faces meeting at their edges, the structure as a whole being characterized in terms of the incidence relations existing between the vertices and edges. Direct mechanisms were obtained from the intersection of the i hyperplanes representing cycles, i.e., closed successions of edges, or pairs of independent mechanisms that produce the same reaction in opposite directions, selected from a possible set of j hyperplanes representing the mechanisms. A computer algorithm screened each of the $\binom{i}{j}$ ways of selecting the i hyperplanes. Any selection yields a system of i linear equations and describes a direct mechanism iff the i hyperplanes intersect at a point. The coefficients of the direct mechanism can be determined by solution of the linear equations. An estimate of the mechanistic complexity of a reaction based on the kinetic laws of chemical reactions has been advanced by Bonchev et al. (1987). Moreover, a comprehensive review and a systematic analysis of the concept of the reaction mechanism as employed in the chemical literature has been presented by Masavetas (1988).

Extensive use has been made of computer programs in recent years for the study of reacting systems. For instance, plausible routes for the synthesis of complex target molecules from comparatively simple starting structures, known as *synthons*, have been devised. The concept of the synthon as a minimal fragment of a molecular system that is necessary for some given chemical reaction to take place was first proposed by Corey (1967). Since that time, the synthon has become an object of mathematical investigation in its own right and various mathematical models of synthons have been put forward. Two examples are those of Koča based on a matrix formulation (Koča 1989) and of Kvasnička and Pospíchal based on a graph-theoretical model (Kvasnička and Pospíchal 1990). The latter work elaborated the concept by introducing additional concepts such as synthon stability, subsynthons and families of isomeric synthons. The new concepts were used in the development of a theory of synthons that closely reflects the reasoning of synthetic chemists. Sets of all possible precursors and successors of given synthons are generated from either linear or cyclic graphs of the relevant reaction mechanism. This means that even chemical reactions having complicated reaction graphs can be expressed as sequences of simple graphs of the reaction mechanism.

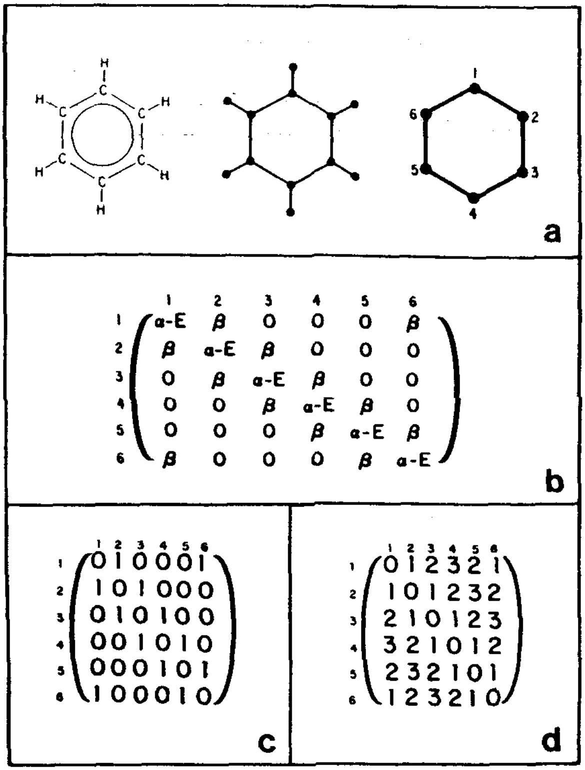
The prediction of plausible routes for chemical syntheses is feasible only because the number of known reaction types is finite and fairly small. Programs that construct reaction sequences to proceed from a small number of synthons to some target molecule usually explore exhaustively every possible combination of routes, and report on those that seem most promising in the probable order of their success. There are, however, currently two fundamentally different approaches to the design of synthesis programs (Ash et al. 1985). One utilizes a library of known reactions and proceeds from the target molecule in a retrosynthetic direction; examples of such programs are LHASA (logic and heuristics applied to synthetic analysis), SYNCHEM (synthetic chemistry), SECS (simulation and evaluation of chemical synthesis) and CASP (computer-aided synthesis program). The other approach is nonempirical in that it attempts to represent structural transformations in a generic way by manipulating an abstract model of the atoms and chemical bonds involved in the reaction sequence. Examples of this approach include EROS (elaboration of reactions for organic synthesis) and CAMEO (computer assisted mechanistic evaluation of organic reactions). Whereas the former approach is essentially a brute force method, the latter is more subtle in that is can search for novel synthesis routes in unexplored areas of chemistry and so make completely new syntheses a reality.

Of these two approaches, the latter is likely to be of greater benefit to the synthetic chemist in the long term. One major advantage is that it harnesses the methodology of artificial intelligence for use in the chemical domain. In particular, it promises the possibility of expert systems that can elicit new chemical knowledge (Pierce and Hohne 1986). The first use of artificial intelligence in synthetic chemistry was made in the late 1960s. The DENDRAL (dendritic algorithm) program was published in 1969 (see Lindsay et al. 1980) for the purpose of generating exhaustive lists of chemically meaningful isomers of structures containing carbon, hydrogen, nitrogen and oxygen atoms. This program used a linear notation for tree structures that started from the center of the tree and avoided nested parentheses. Heuristic DENDRAL is able to construct all of the isomers compatible with experimental data. Since this early work, the applications of artificial intelligence have become increasingly sophisticated. For example, in recent years the expert system SYNLMA (synthesis with logic machine architecture) has used a theorem prover based on a collection of Pascal subroutines in which the target structure is regarded as the theorem to be proved and the starting materials are the axioms (Wang et al. 1986). The expert system SYNCHEM2 is based on a graph embedding technique that determines whether a guest graph can be embedded in a host graph with preservation of adjacency relationships, vertex and edge labeling, and stereochemical orientation. Graph embedding, which is an NP-complete problem, is thus of relevance in the chemical context (Benstock et al. 1988).

5. Polynomials in bonding theory

The (vertex) adjacency matrix of a graph, denoted here as $A(G)$, has been widely used in the modeling of individual chemical molecules and a variety of other chemical systems. Moreover, a fair number of chemical combinatorial problems have arisen from studies on $A(G)$ and the polynomials that may be derived from it. Examples of its uses include the characterization of molecules, the derivation of graph invariants to model chemical systems, and the elucidation of bonding theory. In fig 5.1(c) we illustrate the adjacency matrix for the six atoms forming the carbon skeleton of the benzene molecule (C_6H_6). Because $A(G)$ characterizes graphs up to isomorphism, the matrix has been employed to represent chemical species, including many isomers. In fact, the connection tables used by Chemical Abstracts Service are essentially adjacency matrices in that they list the atoms in a molecule with all the bonding linkages. Over 10 million connection tables representing all known molecules are stored in Chemical Abstracts Service registry files at present. Numerous attempts have been made to transform connection tables into a unique representation; each has sought to give a unique and invariant ordering to the structure. One of the simplest of these was developed by Morgan (see Ash et al. 1985); this partitions atoms into classes based on their connectivity. An iterative process is employed that successively calculates higher-order connectivities for each atom in the molecule. The process terminates when the number of classes ceases to increase. This is the process currently used by Chemical Abstracts Service for the unique labeling of structures in its databases.

As $A(G)$ is not appropriate for the characterization of stereoisomers it has been necessary to modify the Morgan algorithm. Modifications to date are able to accommodate not only the stereochemistry but also additional properties of the atoms bonded in structures, such as their atomic number, and the various kinds of chemical bonding encountered (Ash et al. 1985). For instance, the SEMA (stereochemically extended Morgan algorithm) is able to produce a stereochemi-



$$\begin{array}{c|cccccc} & 1 & 2 & 3 & 4 & 5 & 6 \\ \hline 1 & \alpha\text{-E} & \beta & 0 & 0 & 0 & \beta \\ 2 & \beta & \alpha\text{-E} & \beta & 0 & 0 & 0 \\ 3 & 0 & \beta & \alpha\text{-E} & \beta & 0 & 0 \\ 4 & 0 & 0 & \beta & \alpha\text{-E} & \beta & 0 \\ 5 & 0 & 0 & 0 & \beta & \alpha\text{-E} & \beta \\ 6 & \beta & 0 & 0 & 0 & \beta & \alpha\text{-E} \end{array}$$

$$\begin{array}{c|cccccc} & 1 & 2 & 3 & 4 & 5 & 6 \\ \hline 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 2 & 1 & 0 & 1 & 0 & 0 & 0 \\ 3 & 0 & 1 & 0 & 1 & 0 & 0 \\ 4 & 0 & 0 & 1 & 0 & 1 & 0 \\ 5 & 0 & 0 & 0 & 1 & 0 & 1 \\ 6 & 1 & 0 & 0 & 0 & 1 & 0 \end{array}$$

$$\begin{array}{c|cccccc} & 1 & 2 & 3 & 4 & 5 & 6 \\ \hline 1 & 0 & 1 & 2 & 3 & 2 & 1 \\ 2 & 1 & 0 & 1 & 2 & 3 & 2 \\ 3 & 2 & 1 & 0 & 1 & 2 & 3 \\ 4 & 3 & 2 & 1 & 0 & 1 & 2 \\ 5 & 2 & 3 & 2 & 1 & 0 & 1 \\ 6 & 1 & 2 & 3 & 2 & 1 & 0 \end{array}$$

Figure 5.1. Some graphical and matricial representations of the benzene molecule, C_6H_6. From the left in (a) are depicted the usual chemical structural formula, the graph for benzene, and the graph for the carbon skeleton (C_6) of benzene. (b) shows the Hückel matrix (c) the adjacency matrix, and (d) the distance matrix of the C_6 ring of benzene.

cally unique name for a molecule via a connection table generated from a graphics structure input. This name consists of a variable-length string with components specifying the length of the string, the number of non-hydrogen atoms, bonds including multiple bonds, and tetrahedral stereocenters (see Ash et al. 1985). In the case of constitutional isomers, in which the stereochemistry plays no role, $A(G)$ has frequently been expressed in the form of a code. A unique characterization can be achieved via a canonical numbering of the graph vertices such that the string of binary digits in $A(G)$ read from left to right yields the smallest binary representation. Codes of this type have been employed to establish both the isomorphism of chemical structures and the equivalence of atoms within a given structure by means of computer algorithms (Randić et al. 1981). The derivation of chemically relevant graph invariants from $A(G)$ is addressed in the next section.

nical use of $A(G)$ to date has been in the development of
G) came into prominence in the 1950s when it was realized
phic to the historically important Hückel matrix, $H(G)$ (see
allion 1982). $H(G)$ is a real, symmetric matrix originally derived
. result of the quantum-chemical analysis of hydrocarbon species;
represent the energy levels in molecules and its eigenvectors yield
the electronic charge density within molecules (Mallion 1982). The
Hu ix for the carbon skeleton of the benzene ring is shown in fig. 5.1(b). Before the advent of high-speed computers, the Hückel approach afforded a convenient means of obtaining an approximate solution to the Schrödinger wave equation for the species in question. Although this approach has long since been superseded by more sophisticated methods, it can still provide valuable insights into the behavior of chemical systems and remains a useful pedagogical tool. A number of the results obtained using Hückel theory retain their validity even for the higher approximation methods. For example, in so-called alternant hydrocarbons (which possess bipartite graphs) $A(G)$ assumes the general form,

$$A(G) = \begin{bmatrix} 0 & B \\ B^{t} & 0 \end{bmatrix}, \tag{5.1}$$

where B^{t} is the transpose of the submatrix B. As a consequence of this, the eigenvalues λ_i of such graphs will exist as pairs, i.e.,

$$\lambda_i = -\lambda_{\tau} + 1 - i \quad (1 < i < \tfrac{1}{2}(\tau - 1)) . \tag{5.2}$$

This result, which is of considerable interest to chemists, was first obtained in the chemical context by Coulson and Rushbrooke (1940) and is known to chemists as the pairing theorem (see Mallion and Rouvray 1990). The theorem had however been adumbrated in the work of Perron and Frobenius dating from 1907–1912. They had shown that, if $A(G)$ has h eigenvalues $\lambda_0, \lambda_1, \ldots, \lambda_{h-1}$ viewed as a system of points in the Argand λ-plane, the spectrum maps onto itself under rotation by $2\pi/h$. However, Perron and Frobenius had not made the connecting observation that $A(G)$ for a bipartite graph is an irreducible matrix with $h = 2$, whereas $A(G)$ for a nonbipartite graph is such a matrix with $h = 1$. Had this sequitur been made at the time the mathematical analogue of the Coulson–Rushbrooke theorem could have been proved by 1912. For a discussion of the history of this theorem, see the review of Mallion and Rouvray (1990).

A polynomial derived from $A(G)$, known as the characteristic or spectral polynomial, has been studied by chemists almost as much as $A(G)$ itself. The characteristic polynomial, $P(G)$, is defined by the relation

$$P(G) = \det(xI - A(G)) = \sum_{i=0}^{\tau} a_i x^{\tau - i} , \tag{5.3}$$

where I is the $n \times n$ identity matrix. The roots of $P(G)$ are the eigenvalues of $A(G)$ and these may be ordered into a sequence known as the spectrum of G. Much effort has been expended on determining both the eigenvalues of $A(G)$ and the coefficients a_i of $P(G)$ since the inception of Hückel theory (see Rouvray 1976). Many expressions in closed form have been developed in the chemical context for both the eigenvalues and the coefficients (Graovac et al. 1977, Trinajstić 1983). Expressions for the a_i of graphs of polycyclic aromatic hydrocarbons (represented as polyhexes or hexagonal animal graphs) in terms of graph invariants of these systems have been given by Dias (1985). Among procedures for determining the a_i, mention should be made of Coates' formula, methods based on Ulam subgraphs, recurrence formulas, and the Le Verrier–Faddeev–Frame method; all these have been surveyed by Trinajstić (1988). Coates' formula as used by Sachs, for instance, gives the $a_i(G)$ as

$$a_i(G) = \sum_{s \in S_i} (-1)^{u(s)} 2^{c(s)}\,, \tag{5.4}$$

where s belongs to the set of so-called Sachs graphs S_i on i nodes, and $u(s)$ and $c(s)$ denote respectively the number of components and the number of circuits in s. The components of a Sachs graph are either K_2 graphs or C_p cycles ($p = 3, 4, \ldots, n$) or combinations thereof. One recent example of an efficient computer algorithm developed by a chemist for evaluating $P(G)$ (of the order n^3, where n is the number of graph vertices) based on the Le Verrier–Faddeev–Frame method is that of Živković (1990). The numerous applications of $P(G)$ in chemistry, especially in the context of bonding theory, have been discussed by Trinajstić (1988).

If the C_p cycles are omitted from (5.4), a new polynomial is obtained known to mathematicians as the matching polynomial and to chemists as the acyclic polynomial. The $a_i^{\mathrm{ac}}(G)$ coefficients for this polynomial, designated below as $M(G)$, thus assume the form

$$a_i^{\mathrm{ac}}(G) = \sum_{s \in S_i^{\mathrm{ac}}} (-1)^{u(s)}\,, \tag{5.5}$$

where S_i^{ac} is the set of acyclic Sachs graphs on i nodes. This polynomial has been independently introduced into the literature several times because of its manifold applications in physics and chemistry (Cvetković et al. 1988, Godsil, chapter 31 in this Handbook). One of the major uses has been as a reference polynomial to enable comparisons to be made between the bonding energy in otherwise equivalent molecules which exist in the form of both a chain and a closed ring structure. The difference is known to chemists as the topological resonance energy and affords valuable insights into the stability of molecules (Trinajstić 1983). The polynomial can also be interpreted as a generating function for the number of matchings in graphs. If we denote the number of k-matchings in G as $N(G, k)$, the matching polynomial may be expressed as

$$M(G) = \sum_{k=0}^{[n/2]} (-1)^k N(G, k) x^{n-2k}\,. \tag{5.6}$$

For all graphs G, $N(G, 0)$ is defined to be one, $N(G, 1)$ is equal to the number of edges m and $N(G, k) = 0$ for all $k > m$. For n even, the value of $N(G, n/2)$ is described in chemical terminology as being equal to the number of Kekulé structures K of the molecular graph and the matching is said to be perfect. Kekulé structures have played a major role in the development of chemical bonding theory, especially insofar as this relates to the stability of molecules whose graphs assume the form of polyhexes (Trinajstić 1983). One interesting observation is that the number of Kekulé structures is closely related to the terminal coefficients of $P(G)$ and $M(G)$, i.e.,

$$a_n(G) = K^2 \tag{5.7}$$

and

$$a_n^{\mathrm{ac}}(G) = K\,. \tag{5.8}$$

Further analysis of these polynomials in the chemical context, the development of a more general polynomial incorporating both $P(G)$ and $M(G)$ as special cases, and discussion of a number of related polynomials are to be found in the monographs of Gutman and Polansky (1986) and Cvetković et al. (1988).

6. Invariants and molecular properties

In addition to polynomials, chemists have made extensive use of scalar numerical graph invariants for the characterization of chemical species. The first such invariant was introduced some 150 years ago when the number of carbon atoms in hydrocarbon molecules was adopted for the characterization of these species (Rouvray 1990a). This particular invariant is known today as the carbon number index because it equals the number of carbon atoms in the hydrogen-deleted molecular graph of a hydrocarbon species. (Deletion of the hydrogen atoms is a common practice in this field as the hydrogen atoms are usually not structure-determining.) Scalar numerical invariants are nowadays commonly referred to by chemists as *topological indices*. Another example of such an invariant is the cyclomatic number (see section 2), which has been in use for the characterization of molecules for well over a century (see Rouvray 1975). The chemical fascination with such invariants stems from the fact that they can be interpreted as a property of a molecule on a par with measured properties such as the reactivity. In fact, once a correlation with an invariant has been established for a test subset of related molecules, prediction of the property in question can normally be made for all the remaining molecules in the set (see Rouvray 1986).

In many instances, the use of numerical invariants for property prediction rests on an implicit application of the additivity principle, which holds whenever a molecular property may be obtained as the sum of the contributions to that property from each of the constituent parts of the molecule (See Rouvray 1990a). More formally, a molecular property θ of some systems $\mathcal{S}$ is said to be additive whenever that system can be broken down into noninteracting subsystems ω_1 and

ω_2 according to the equation

$$\theta(\mathcal{S}) \rightarrow \theta(\omega_1 \cup \omega_2) \rightarrow \theta(\omega_1) + \theta(\omega_2) . \tag{6.1}$$

Such properties are often encountered when sets of related molecules are considered, e.g., the members of homologous series. Other types of systematic interaction have also been explored and formulated in mathematical language by Klein (1986). For the additive case, Gordon and Kennedy (1973) showed that any molecular property can be expressed as a linear combination of graph-theoretical invariants thus:

$$\theta(\mathcal{S}) = \sum_i k_i \xi_i , \tag{6.2}$$

where the ξ_i are scalar numerical invariants derived from chemical graphs and k_i are empirically determined coefficients. Properties satisfying this type of relationship are said to belong to the "graph-like state of matter" (Gordon and Kennedy 1973). Examples of properties that approximate to this state are the energy, entropy, melting and boiling points, and the refractive index.

To date, over 150 different ξ_i graph invariants have been put forward in the chemical literature. Just under one half of these are purely graph-theoretical and the remainder are information-theoretical in nature (see Bonchev 1983). There are numerous reviews now available on these invariants – see, for example, Bonchev (1983), Rouvray (1985, 1986, 1987, 1989), Kier and Hall (1986), and Stankevich et al. (1988). A fair number of these invariants derive from either the adjacency matrix $A(G)$ or the distance matrix $D(G)$ of the molecular graph (Rouvray 1985). The distance matrix for the carbon skeleton of the benzene molecule is illustrated in fig. 5.1(d). The first of the invariants designed to characterize branched molecules was introduced in 1947 and is referred to today as the Wiener index (see Rouvray 1985). This index is defined by the equation

$$W(G) = \tfrac{1}{2} \sum_i \sum_j d_{ij}(G) , \tag{6.3}$$

i.e., the index is half the sum of the elements of $D(G)$. Expressions for $W(G)$ in closed form are known for many different classes of graph. For instance, the form for a path graph is $\frac{1}{6}(n^3 - n)$ and that for a star graph is $(n-1)^2$. A closed expression for the general tree has been developed by Canfield et al. (1985). The Wiener index has found application in numerous areas of chemistry and physics, and is now regarded as one of the most successful invariants for the prediction of physicochemical properties (Rouvray 1985).

One of the most interesting invariants from the mathematical point of view is that of Hosoya published in 1971 (Hosoya 1985). This invariant he described as the *topological index*, a term that has since been extended to embrace all scalar numerical graphs invariants. The Hosoya index is defined as follows:

$$H(G) = \sum_{k=0}^{[n/2]} N(G, k) , \tag{6.4}$$

where $N(G, k)$ is the number of k-matchings in G referred to in section 5. $H(G)$ was adapted from earlier work in statistical physics in which matchings and polynomials were (and still are) employed as a means of modeling the coverings of crystal lattices by diatomic molecules or so-called dimers. A history of this earlier work together with a discussion on the physical applications of the matching polynomial $M(G)$ are to be found in the work of Heilmann and Lieb (1972). It is thus only to be expected that $H(G)$ will be related to several other graph invariants, including a number of polynomials. For instance, the values assumed by $H(G)$ for path graphs form members of the Fibonacci series, and the values for monocycles form members of the Lucas series (see Rouvray 1987). There are now extensive tabulations available for both $H(G)$ and $N(G, k)$ values (Hosoya 1985). $H(G)$ is closely associated with the characteristic polynomial $P(G)$ of a graph G. In the case of a tree T, the relationship takes the form

$$P_T(x) = \sum_{k=0}^{[n/2]} (-1)^k N(T, k) x^{n-2k} . \tag{6.5}$$

For cycles, additional terms need to be added on the right-hand side (Hosoya 1985). Moreover, the counting polynomials

$$Q_G(x) = \sum_{k=0}^{[n/2]} N(T, k) x^k \tag{6.6}$$

for various classes of graphs G are known to transform into a family of orthogonal polynomials, including the first and second types of Chebyshev, Hermite, Laguerre and associated Laguerre polynomials (Hosoya 1985), see Heilmann and Lieb (1972) for Chebyshev and Hermite polynomials. The chemical applications of $H(G)$ have been summarized by Rouvray (1987).

The most popular of the invariants in terms of the number of applications to date is the so-called molecular connectivity index of Randić first proposed in 1975 (see Kier and Hall 1986). Two entire monographs have been devoted to exposition of this index and its various successors (Kier and Hall 1986). In its original form, the index was defined as follows:

$$\chi(G) = \sum_{\text{edges}} (\delta_i \delta_j)^{-1/2} \tag{6.7}$$

where the δ_i and δ_j represent the degrees of an adjacent pair of vertices i and j in G. The index was subsequently generalized into a series of indices in which the summation is made over a variety of subgraphs of G other than the edges. In its most general form, $\chi(G)$ for tree graphs may be expressed as

$$h_{\chi_r}(G) = \sum_{k=1}^{\sigma_h} \prod_{i=1}^{h+1} (\delta_i)_k^{-1/2} , \tag{6.8}$$

where h is the number of edges in the subgraph used in the summation, r is the type of subgraph used, and σ_h is the number of subgraphs of type r with h edges.

So far, very little mathematical analysis of this index has been made, though it is evident that $h_\chi(G)$ can be derived from the hth power of $A(G)$. The role of the index in characterizing branching has also been discussed (Kier and Hall 1986, Rouvray 1988). The index has been employed in successful correlations with a host of molecular properties ranging from the physical, through the chemical, to the biological (Kier and Hall 1986).

7. Some outstanding problems

In this section brief reference is made to some of the many important mathematical problems that still need to be solved in chemical combinatorics. Many of these problems are of interest to chemists and combinatorialists alike. Examples of such problems include the general solution of the hexagonal animal enumeration problem (which is equivalent to enumerating polycyclic aromatic hydrocarbons), and the characterization of the eigenvalue spectra, especially the occurrence of degeneracies, of the many polynomials used in bonding theory and other chemical applications. Such polynomials include the characteristic polynomial, the matching polynomial, the chromatic polynomial, the distance polynomial, random walk counting polynomials, the sextet polynomial, the permanental polynomial, and the polynomial associated with the Ising model partition function (Trinajstić et al. 1986). These and other outstanding problems have been described and listed by Rouvray (1985) and Trinajstić et al. (1986). The challenges of characterizing branching in molecular species (Rouvray 1988), the computer perception of molecular symmetry (Randić et al. 1981), and the description of molecular similarity based on metric spaces (Herndon and Bertz 1987, Johnson and Maggiora 1990) are examples of other problems that need to be addressed. It would seem with all these problems (and the many others not mentioned here because of space limitations) that both chemists and combinatorialists will have more than enough to keep them occupied for at least another century. This brings to mind the words of the mathematician Sylvester first stated in 1878: "There is a wealth of untapped mathematical potential contained in the patient and long investigations of our chemist fellows".

References

Ash, J.E., P.A. Chubb, S.E. Ward, S.M. Welford and P. Willet
[1985] *Communication, Storage and Retrieval of Chemical Information* (Ellis Horwood, Chichester).

Balaban, A.T.
[1976] ed., *Chemical Applications of Graph Theory* (Academic Press, London).
[1985] Applications of graph theory in chemistry, *J. Chem. Inform. Comput. Sci.* **25**, 334–343.
[1987] *Annulenes, Benzo-, Hetero-, Homo-Derivatives and their Valence Isomers* (CRC Press, Boca Raton, FL).

Balasubramanian, K.
[1985] Applications of combinatorics and graph theory to spectroscopy and quantum chemistry, *Chem. Rev.* **85**, 599–618.

Benstock, J.D., D.J. Berndt and K.K. Agarwal
[1988] Graph embedding in SYNCHEM2, an expert system for organic synthesis discovery, *Discrete Appl. Math.* **19**, 45-63.

Bonchev, D.
[1983] *Information Theoretic Indices for Characterization of Chemical Structures* (Research Studies Press, Letchworth, England).

Bonchev, D., and D.H. Rouvray, eds.
[1991] *Chemical Graph Theory: Introduction and Fundamentals* (Abacus/Gordon and Breach, London).
[1992] *Chemical Graph Theory: Reactivity and Kinetics* (Abacus/Gordon and Breach, London).

Bonchev, D., D. Ramensi and O.N. Temkin
[1987] Complexity index for the linear mechanisms of chemical reactions, *J. Math. Chem.* **1**, 345–388.

Brocas, J.
[1986] Double cosets and enumeration of permutational isomers of fixed symmetry, *J. Amer. Chem. Soc.* **108**, 1135–1145.

Brocas, J., M. Gielen and R. Willem
[1984] *The Permutational Approach to Dynamic Stereochemistry* (McGraw-Hill, New York).

Canfield, E.R., R.W. Robinson and D.H. Rouvray
[1985] Determination of the Wiener molecular branching index for the general tree, *J Comput. Chem.* **6**, 598-609.

Cayley, A.
[1875] On the analytical forms called trees, with application to the theory of chemical combinations, *Rep. Brit. Assoc. Adv. Sci.* **45**, 257–305.

Corey, E.J.
[1967] General methods for the construction of complex molecules, *Pure Appl. Chem.* **14**, 19–37.

Coulson, C.A., and G.S. Rushbrooke
[1940] Note on the method of molecular orbitals, *Proc. Cambridge Philos. Soc.* **36**, 193 200.

Crum Brown, A.
[1865] On the use of graphic representations of chemical formula, *Proc. R. Soc. Edinburgh* **5**, 429 430.

Cvetković, D.M., M. Doob, I. Gutman and A. Torgasev
[1988] *Recent Results in the Theory of Graph Spectra* (North-Holland, Amsterdam).

Davidson, R.A.
[1981] Isomers and isomerizations: elements of Redfield's combinatorial theory, *J. Amer. Chem. Soc.* **103**, 312–314.

Dias, J.R.
[1985] Properties and derivation of the fourth and sixth coefficients of the characteristic polynomial of molecular graphs, *Theor. Chim. Acta* **68**, 107-123.
[1987] *Handbook of Polycyclic Hydrocarbons, Part A* (Elsevier, Amsterdam).

Golender, V.E., and A.B. Rozenblit
[1983] *Logical and Combinatorial Algorithms for Drug Design* (Research Studies Press, Letchworth, England).

Gordon, M., and J.W. Kennedy
[1973] The graph-like state of matter. II. LCGI schemes for the thermodynamics of alkanes and the theory of inductive inference, *J. Chem. Soc. Faraday Trans. II* **69**, 484–504.

Graovac, A., I. Gutman and N. Trinajstić
[1977] *Topological Approach to the Chemistry of Conjugated Molecules, Lecture Notes in Chemistry,* Vol. 4 (Springer, Berlin).

Gutman, I., and O.E. Polansky
[1986] *Mathematical Concepts in Organic Chemistry* (Springer, Berlin).

Hässelbarth, W.
[1987] The inverse problem of isomer enumeration, *J. Comput. Chem.* **8**, 700 717.

Heilmann, O.J., and E.H. Lieb
[1972] Theory of monomer dimer systems, *Comm. Math. Phys.* **25**, 190 232.

Herndon, W.C., and S.H. Bertz
[1987] Linear notations and molecular graph similarity, *J. Comput. Chem.* **8**, 367–374.

Hosoya, H.
[1985] Topological index as a common tool for quantum chemistry, statistical mechanics, and graph theory, in: *Mathematics and Computational Concepts in Chemistry*, ed. N. Trinajstić (Ellis Horwood, Chichester), pp. 110–123.

Johnson, M.A., and G.M. Maggiora
[1990] eds., *Concepts and Applications of Molecular Similarity* (Wiley-Interscience, New York).

Kennedy, J.W., and L.V. Quintas
[1988] eds., *Applications of Graphs in Chemistry and Physics* (North-Holland, Amsterdam).

Kier, L.B., and L.H. Hall
[1986] *Molecular Connectivity in Structure–Activity Analysis* (Research Studies Press, Letchworth, England).

King, R.B.
[1969] Chemical applications of topology and group theory. I. Coordination polyhedra, *J. Amer. Chem. Soc.* **91**, 7211 7216.
[1983] ed., *Chemical Applications of Topology and Graph Theory* (Elsevier, Amsterdam).
[1988] Topological aspects of polyhedral isomerizations, in: *Advances in Dynamic Stereochemistry*, ed. M. Gielen (Freund Publishing House, Tel Aviv, Israel) pp. 1–36.

King, R.B., and D.H. Rouvray
[1987] eds., *Graph Theory and Topology in Chemistry* (Elsevier, Amsterdam).

Klein, D.J.
[1986] Chemical graph-theoretic cluster expansions, *Int. J. Quant. Chem., Quant. Chem. Symp.* **20**, 153–171.

Klein, D.J., and M. Randić
[1990] eds., *Methods of Mathematical Chemistry* (Baltzer, Basel).

Klemperer, W.G.
[1972] Enumeration of permutational isomerization reactions, *J. Chem. Phys.* 56, 5478–5489.

Knop, J.V., W.R. Müller, K. Szymanski and N. Trinajstić
[1985] *Computer Generation of Certain Classes of Molecules* (Assoc. Chem. Technol. of Croatia, Zagreb, Yugoslavia).
[1990] Use of small computers for large calculations: Enumeration of polyhex hydrocarbons, *J. Chem. Inform. Comput. Sci.* **30**, 159–160.

Koča, J.
[1989] A mathematical model of realistic constitutional chemistry. A synthon approach. I. An algebraic model of synthon, *J. Math. Chem.* **3**, 73–89.

Lindsay, R.K., B.G. Buchanan, E.A. Feigenbaum and J. Lederberg
[1980] *Applications of Artificial Intelligence for Organic Chemistry* (McGraw-Hill, New York).

Lloyd, E.K.
[1988] Redfield's papers and their relevance to counting isomers and isomerizations, *Discrete Appl. Math.* **19**, 289–304.

Mallion, R.B.
[1982] Some chemical applications of the eigenvalues and eigenvectors of certain finite, planar graphs, in: *Applications of Combinatorics*, ed. R.J. Wilson (Shiva Publishing, Nantwich, Cheshire, England) pp. 87–114.

Mallion, R.B., and D.H. Rouvray
[1990] The golden jubilee of the Coulson–Rushbrooke Pairing Theorem, *J. Math. Chem.* 5, 1–21.

Maruani, J., and J. Serre
[1983] eds., *Symmetries and Properties of Non-Rigid Molecules* (Elsevier, Amsterdam).

Masavetas, K.A.
[1988] Mathematical properties common in all mechanism models of chemical reactions, *Math. Comput. Model.* **10**, 263 274.

Merrifield, R.E., and H.E. Simmons
[1989] *Topological Methods in Chemistry* (Wiley-Interscience, New York).

Milner, P.C.

[1964] The possible mechanisms of complex reactions involving consecutive steps, *J. Electrochem. Soc. III*, pp. 228–232.

Pierce, T.H., and B.A. Honne

[1986] eds., *Artificial Intelligence Applications in Chemistry* (American Chemical Society, Washington, DC).

Pólya, G.

[1937] Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen, *Acta Math.* **68**, 145–254.

Pólya, G., and R.C. Read

[1987] *Combinatorial Enumeration of Groups, Graphs and Chemical Compounds* (Springer, New York).

Pospichal, J., and V. Kvasnička

[1990] Graph theory of synthons, *Int. J. Quant. Chem.* **38**, 253–278.

Randić, M., G.M. Brissey and C.L. Wilkins

[1981] Computer perception of topological symmetry via canonical numbering of atoms, *J. Chem. Inform. Comput. Sci.* **21**, 52–59.

Redfield, J.H.

[1927] The theory of group-reduced distributions, *Amer. J. Math.* **49**, 433–455.

Rouvray, D.H.

[1974] Isomer enumeration methods, *Chem. Soc. Rev. (London)* **3**, 355–372.

[1975] Some reflections on the topological structure of covalent molecules, *J. Chem. Educ.* **52**, 768–773.

[1976] The topological matrix in quantum chemistry, in: *Chemical Applications of Graph Theory*, ed. A.T. Balaban (Academic Press, London) pp. 175–221.

[1985] The role of the topological distance matrix in chemistry, in: *Mathematics and Computational Concepts in Chemistry*, ed. N. Trinajstić (Ellis Horwood, Chichester), pp. 295–306.

[1986] Predicting chemistry from topology, *Sci. Amer.* **254**(9), 40–47.

[1987] The modeling of chemical phenomena using topological indices, *J. Comput. Chem.* **8**, 470–480.

[1988] The challenge of characterizing branching in molecular species, *Discrete Appl. Math.* **19**, 317–338.

[1989] The limits of applicability of topological indices, *J. Mol. Struct. Theochem* **185**, 187–201.

[1990a] The origins of chemical graph theory, in: *Mathematical Chemistry*, Vol. 1, eds. D. Bonchev and D.H. Rouvray (Gordon and Breach, London) pp. 1–30.

[1990b] ed., *Computational Chemical Graph Theory* (Nova Science, New York).

Rouvray, D.H., and A.T. Balaban

[1979] Chemical applications of graph theory, in: *Applications of Graph Theory*, eds. R.J. Wilson and L.W. Beineke (Academic Press, London) pp. 177–221.

Ruch, E., and D.J. Klein

[1983] Double cosets in chemistry and physics, *Theor. Chim. Acta* **63**, 447–472.

Ruch, E., W. Hässelbarth and B. Richter

[1970] Doppelnebenklassen als Klassenbegriff und Nomenklaturprinzip für Isomere und ihre Abzahlung, *Theor. Chim. Acta* **19**, 288–300.

Sellers, P.H.

[1967] Algebraic complexes which characterize chemical networks, *SIAM J. Appl. Math.* **15**, 13–68.

[1984] Combinatorial classification of chemical mechanisms, *SIAM J. Appl. Math.* **44**, 784–792.

[1989] Combinatorial aspects of enzyme kinetics, in: *Applications of Combinatorics and Graph Theory to the Biological and Social Sciences*, ed. F. Roberts (Springer, New York) pp. 295–314.

Simon, J.

[1987] A topological approach to the stereochemistry of nonrigid molecules, in: *Graph Theory and Topology in Chemistry*, eds. R.B. King and D.H. Rouvray (Elsevier, Amsterdam) pp. 43–75.

Slanina, Z.

[1986] *Contemporary Theory of Chemical Isomerism* (Reidel, Dordrecht).

Stankevich, M.I., I.V. Stankevich and N.S. Zefirov

[1988] Topological indices in organic chemistry, *Russ. Chem. Rev.* **57**, 191–208.

Sylvester, J.J.
[1878] On an application of the new atomic theory to the graphical representation of the invariants and covariants of binary quantics, *Amer. J. Math.* **1**, 64-125.

Trinajstić, N.
[1983] *Chemical Graph Theory,* two volumes (CRC Press, Boca Raton, FL); Second edition: 1992, one volume.
[1988] The characteristic polynomial of a chemical graph, *J. Math. Chem.* **2**, 197–215.

Trinajstić, N., D.J. Klein and M. Randić
[1986] On some solved and unsolved problems of chemical graph theory, *Int. J. Quant. Chem., Quant. Chem. Symp.* **20**, 699–742.

Wang, T., I. Burnstein, M. Corbett, S. Ehrlich, M. Evens, A. Gough and P. Johnson
[1986] Using a theorem prover in the design of organic synthesis, in: *Artificial Intelligence Applications in Chemistry,* eds. T.H. Pierce and B.A. Hohne (American Chemical Society, Washington, DC).

Živković, T.P.
[1990] On the evaluation of the characteristic polynomial of a chemical graph, *J. Comput. Chem.* **11**, 217 222.

CHAPTER 39

Applications of Combinatorics to Molecular Biology

Michael S. WATERMAN

Departments of Mathematics and Molecular Biology, University of Southern California, Los Angeles, CA 90089-1113, USA

Contents

This work was supported by grants from the National Institutes of Health, the National Science Foundation and the System Development Foundation.

HANDBOOK OF COMBINATORICS
Edited by R. Graham, M. Grötschel and L. Lovász

1. Introduction

Combinatorics in molecular biology 1985 The biological sciences have undergone a revolution in the last dozen years. Al-most every edition of a major newspaper reports some new discovery in biology, often with medical and/or financial implications. Biologists now have the ability to rapidly read and manipulate DNA, the basic material of life that makes up chromosomes and is the carrier of genetic information. The reading of DNA is called sequencing, since the scientists are determining the linear sequence of bases along the DNA molecule. The bases or alphabet of DNA is adenine (/t), guanine (G), cytosine (C), and thymine (T). These bases, joined to a sugar-phosphate backbone, are linked together in a chain to form DNA. Frederick Sanger and Walter Gilbert independently developed procedures for the rapid sequencing of long segments of DNA molecules. They received the Nobel Prize in 1980 for their discoveries. Sanger, incidentally, was earlier the first to determine the amino acid sequence of a protein, insulin.

Rapid DNA sequencing has caused an information explosion. It was only in 1953 that a complementary double-helical structure was postulated for DNA. By 1975 only a few hundred bases had been sequenced. In Spring 1994 DNA sequences are collected in international databases and sequences totaling about 200 million bases are known. These sequences come from various locations in the genomes of a wide variety of organisms. (A genome holds all the genetic information of an organism.) The sequences vary greatly in size. A long continuous sequence that has been determined to date is that of human cytomegalovirus which is 229354 bases long.

Early in this century, Fisher, Haldane and Wright did fundamental work in proving that the Mendelian model of genetics, with discrete alleles, is rich enough to generate the seemingly continuous range of phenotypes observed in nature. This might seem almost trivial in light of today’s emphasis on discrete mathematics, but it was by no means obvious at that time. Their mathematical work in population biology led experimental biology. Today mathematical scientists lag far behind the experimental biologists as they read the basic material of the gene and directly test hypotheses about the nature of life. There has developed a small field of mathematical and computer scicnccs to assist the molccular biologist in his endeavor. Most of this mathematical development is about discrete structures. See Waterman (1989).

Increasing attention is being given to the mathematical and computational aspects of molecular biology because of the human genome project. This project can be viewed as directed toward sequencing all the DNA of humans and other organisms. While 2(H) million bases of DNA have been sequenced in pieces that average about 1000 bases long, the genome of even the bacterium E. coli is about 5 million bases. Man has a genome of 3 billion bases. Presently, the efforts center on improving the mapping and sequencing technology so that such sequencing projects can be more easily accomplished. Even so, using today’s technology, genomes of the size of those of E. coli will be sequenced within the next few years. A number of analytical problems are concerning people who are involved in these studies.

First of all, the puzzles of assembling map and sequence information from the experimental results are large, combinatorial problems. In addition, the vast quantity of data will severely tax our current methods for finding the relationships between the sequences that are determined

In this chapter some combinatorial aspects of molecular biology will be explored. Section 2 discusses sequence alignments where certain sequence relationships are studied, both by enumeration and algorithms. The next section gives some results on enumeration and algorithms for secondary structures. The final section treats restriction maps of DNA and their relationship with the human genome project. The emphasis is on the description and straightforward solution of some of the related problems. Recently this general area has become increasingly active.

2. Sequence alignments

Evolution is a key concept in biology. To understand living organisms, biologists study the relationships between the organisms and their environments. Important inventions, such as the eye, are maintained and improved on throughout history. When these concerns of understanding the how and why of biology are brought to a molecular level, the evolutionary mode of thinking is extremely important. Certain machinery such as that involved in DNA to protein translation (the genetic code) is present in all organisms and works everywhere in essentially the same way. These mechanisms are so basic to life and so much additional biological activity depends on them that they cannot be modified except in very minor ways.

Other more recent 'inventions' at the molecular level allow us to understand the difference between life forms in terms of their history. For example, organisms with a nucleus (such as humans) are classified as eukaryotes while those without a nucleus (such as *E. coli*) are classified as prokaryotes. Finer and finer distinctions can be made, and classifying organisms goes hand in hand with understanding how they function.

Something of the same approach is taken by biologists in performing DNA sequence analysis. Given a sequence x, what known sequences are related to it and what are the relationships? Before this question can be explored we need to understand what evolutionary events can take place during sequence evolution. The simplest event is substitution, where one nucleotide is replaced by another, as when A is replaced by C for example. Nucleotides can be inserted into or deleted from a sequence, either one nucleotide at a time or in blocks. Insertions and deletions greatly complicate the analysis. Inversions and duplications of a block of sequence make things even more difficult.

It is common in molecular biology to try to discover the function of a DNA or protein sequence by relating it to other sequences. Frequently this means a biologist will compare a sequence with a large number of previously analyzed sequences; the comparison is done using a computer using algorithms as described below. These comparisons are usually done with sequences taken two at a time. Often there are families of related sequences where any pair might have a fairly weak relationship. Therefore there is a good deal of interest in comparison of more than two sequences, often in comparison of several hundred sequences.

In most sequence analysis the sequence transformations are restricted to substitutions, insertions or deletions. The biologist represents his findings in an alignment of one sequence written over another, and the sequence transformations can be read from the alignment. For example,

$$\begin{array}{l} ATTA-CGG \\ -CGACC-G \end{array}$$

is an alignment of $ATTACGG$ with $CGACCG$. From the point of view of taking the top sequence as the "original" sequence, this alignment shows the events in an evolution of $\boldsymbol{x}$ to y. There has been the deletion of an A and a G, the substitution of C and G for the two T's, and the insertion of a C. There is no history recorded in an alignment, since there is no information about the timing of the events relative to one another nor is it known which sequence "came first". In fact, some other sequence is likely to have been the ancestor of both sequences. In the next section we consider some combinatorics motivated by considering the history of the events, then we discuss sequence alignment combinatorics and algorithms.

2.1. *Shuffles and alignments*

Let $\boldsymbol{x} = x_1x_2\cdots x_n$ and $\boldsymbol{y} = y_1y_2\cdots y_m$ be two sequences. The problem under consideration here is to count the histories for a special type of evolution: delete all the letters of $\boldsymbol{x}$ and insert all the letters of y. The deletion/insertion events take place one letter at a time, the events can be performed in any order and it is possible to track each nucleotide. Thus for simplicity it is assumed that all $n+m$ letters are distinct. The results described in this section are from Greene (1988) where material of independent combinatorial interest also appears. While this is a very special case of molecular evolution, the possible histories between two sequences are of much biological interest. Greene's work is the first mathematical study of this complex problem.

Define an order by $\boldsymbol{s} \leqslant \boldsymbol{t}$ if $\boldsymbol{s}$ is a subsequence of $\boldsymbol{t}$. Let $\{\boldsymbol{s}\}$ denote the set of letters in the sequence $\boldsymbol{s}$, and $\boldsymbol{s}|\boldsymbol{t}$ denote the sequence $\boldsymbol{s}$ restricted to the set $\{\boldsymbol{t}\}$. A sequence $\boldsymbol{s}$ is on a path between $\boldsymbol{x}$ and $\boldsymbol{y}$ if $\{\boldsymbol{s}\} \subset \{\boldsymbol{x}\} \cup \{\boldsymbol{y}\}$ with $\boldsymbol{s}|\boldsymbol{x} \leqslant \boldsymbol{x}$ and $\boldsymbol{s}|\boldsymbol{y} \leqslant \boldsymbol{y}$. This set of sequences is denoted $W(\boldsymbol{x},\boldsymbol{y})$ and was noted by Greene to be shuffles of subsequences of $\boldsymbol{x}$ and $\boldsymbol{y}$. If we maintain the idea of going "from" $\boldsymbol{x}$ "to" $\boldsymbol{y}$, there is a natural partial order on $W(\boldsymbol{x},\boldsymbol{y})$:

$$\boldsymbol{s} \leqslant_* \boldsymbol{t} \text{ if } \begin{cases} \boldsymbol{s}|\boldsymbol{x} \geqslant \boldsymbol{t}|\boldsymbol{x}, \\ \boldsymbol{s}|\boldsymbol{y} \leqslant \boldsymbol{t}|\boldsymbol{y}, \\ \boldsymbol{s}|\boldsymbol{t} = \boldsymbol{t}|\boldsymbol{s}. \end{cases}$$

All sequences of the same length have the same order structure so we set $W(\boldsymbol{x},\boldsymbol{y}) = W(n,m)$. For any n and m, $W(n,m)$ is a lattice.

There are some natural combinatorial questions about $W(n,m)$ such as determining $\Omega(n,m)$, the number of elements. Greene answers virtually all of these questions. We set $C_{n,m}$ equal to the number of maximal chains in $W(n,m)$ and define

$$\Phi_{n,m}(x)=\sum_{j\geqslant 0}\binom{m}{j}\binom{n}{j}x^j.$$

This last function is closely related to the Jacobi polynomials and both of the quantities of interest can be expressed in terms of it.

$$\Omega_{n,m}=2^{n+m}\Phi_{n,m}(1/4),$$
$$C_{n,m}=(n+m)!\Phi_{n,m}(1/2),$$

Many interesting cases remain to be studied. Simply changing one sequence to another by deleting all letters of one sequence and inserting all letters of another is not realistic biology. The extension of Greene's work to allow matching and mismatching letters remains to be made; it is likely to be extremely difficult.

2.2. *Sequence alignment*

In this section we will consider alignment of $\boldsymbol{x}=x_1x_2\cdots x_n$ and $\boldsymbol{y}=y_1y_2\cdots y_n$. The sequences are the same length to avoid non-essential complications of the results. Both algorithms and combinatorics for alignment are easy if no insertions and deletions are allowed. There are simply $2n+1$ ways to align the sequences, one over the other. Each of these alignments can be evaluated for quality of matching by direct examination of the overlapping portions. Therefore the best alignments can be found in $O(n^2)$ time. While this chapter is not intended to be a survey of algorithms for alignments, this area is discussed as it is of great importance in biology. Also, it motivates some useful combinatorics. Insertions and deletions can be included in sequence alignments and best alignments can still be located in $O(n^2)$ time. Reviews of the field have appeared in Kruskal and Sankoff (1983) and in Waterman (1984, 1989).

An alignment can be viewed as a way to extend the sequences to be of the same length L, equal to the overall length of the alignment. The alignment shown above

$$\begin{matrix}ATTA-CGG\\ -CGACC-G\end{matrix}$$

has length $L=8$. Note that the alphabet for the extended sequences has been increased by the symbol "–".

We now turn to asymptotics for the number of alignments of two sequences of length n. The first results for this problem related the number of alignments to the Stanton–Cowan numbers (Laquer 1981). One way to count alignments is to identify aligned pairs $\binom{x_i}{y_j}$ and simply to choose subsets of $\boldsymbol{x}$ and $\boldsymbol{y}$ to align. This gives

$$\sum_{k\geqslant 0}\binom{n}{k}\binom{m}{k}=\binom{n+m}{n}$$

alignments if $\boldsymbol{x}$ has n letters and $\boldsymbol{y}$ has m letters. Recent work has generalized these results. Biologists find an alignment more convincing when the matched segments, that is segments without insertions or deletions, occur in larger blocks. Let $g(b,n)$ be the number of alignments where the matched sections are of length at least b. The following appears in Griggs et al. (1986):

For $b \geqslant 1$ define

$$h(x) = (1-x)^2 - 4x(x^b - x + 1)^2$$

and let $\rho = \min\{x\colon h(x) = 0\}$. Then

$$g(b,n) \sim (\gamma_b n^{-1/2})\rho^{-n} \quad \text{as } n \to \infty,$$

where $\gamma_b = (\rho^b - \rho + 1)(-\pi\rho h'(\rho))^{-1/2}$. The proof uses generating functions for $g(b,n)$. We remark that the result of Laquer (1981) is given by the above result with $b = 1$.

Next are some results on $f(k,n)$, the number of alignments of k sequences of length n (Griggs et al. 1990). Using combinatorial argument to give the exponential growth rate:

For fixed $k \geqslant 2$,

$$\lim_{n\to\infty} \ln(f(k,n))/n = \ln(c_k),$$

where $c_k = (2^{1/k} - 1)^{-k}$. It is also possible to show that the asymptotic behavior of c_k is equivalent to that of $2^{-1/2}(\ln 2)^{-k}k^k$.

Employing a saddle point method gives more precise asymptotics for $f(k,n)$. For fixed $k \geqslant 2$ let $r = (2^{1/k} - 1)^k$. Then

$$f(k,n) = [r^{-n} n^{-(k-1)/2}]\left[(r^k \pi^{(k-1)/2} k^{1/2})^{-1} 2^{(k^2-1)/2k} + \mathrm{O}(n^{-1/2})\right].$$

From the asymptotics given here it is clear that it is not possible to just look at all possible sequence alignments and pick the preferred ones. It is necessary to define an objective function for "good" alignments. Suppose a function $s(a,b)$ is given to score the alignment of a and b from the sequence alphabet, and that the problem is to find the highest scoring alignments. This score is given by

$$S(\boldsymbol{x},\boldsymbol{y}) = \max_{\text{all alignments}} \sum_{1\leqslant j\leqslant L} s(x_j^*, y_j^*),$$

where x_j^* and y_j^* are the jth members of the extended sequences.

A simple dynamic programming method can be used to find the maximum scoring alignment.

Let $\boldsymbol{x} = x_1x_2\cdots x_n$ and $\boldsymbol{y} = y_1y_2\cdots y_n$. Set $S_{0,j} = \sum_{1\leqslant k\leqslant j} s(-,y_k)$, $S_{0,0} = 0$, $S_{i,0} = \sum_{1\leqslant k\leqslant i} s(x_k,-)$, and $S_{i,j} = S(x_1x_2\cdots x_i, y_1y_2\cdots y_j)$. Then $S(\boldsymbol{x},\boldsymbol{y}) = S_{n,n}$ and

$$S_{i,j} = \max\begin{cases} S_{i-1,j} + s(x_i,-), \\ S_{i-1,j-1} + s(x_i,y_j), \\ S_{i,j-1} + s(-,y_j). \end{cases}$$

The above algorithm aligns two sequences in $O(n^2)$ time and space. Letters are inserted or deleted in blocks in biology. For general weighting of these "gaps" the dynamic programming algorithm has time $O(n^3)$ (Waterman 1986), while linear weighting retains time $O(n^2)$. It can be argued that the weighting should be concave where the comparisons can be made in almost the same time. See Waterman (1989), Miller and Myers (1988) and Galil and Giancarlo (1989).

For the case of k sequences of length n the simple algorithm generalizes to require $O(2^k n^k)$ time and space. This is computationally impractical and several different approaches have been taken to solve this important problem; see Waterman (1986) and Waterman and Jones (1990). Some recent approaches to this important problem are now described.

Carrillo and Lipman (1988) consider the generalization of dynamic programming alignments to k sequences. They observe that the score of the projection of a multiple alignment onto two of the sequences cannot be more than the score of those two sequences aligned by themselves. They exploit this observation to greatly reduce the time and storage of multiple sequence alignment. As many as 9 or 10 sequences might be aligned by their technique.

Another approach to k-sequence alignment is to build up the multiple alignment from two sequence alignments. It is obviously possible to begin with the best-aligned sequence pairs and obtain an unsatisfactory result in the end, but some groups have made useful algorithms based on this approach. In Waterman and Perlwitz (1984) some connections with geometry are explored. Taylor (1987) and Vingron and Argos (1989) have excellent programs along these general lines.

Finally in Waterman (1986) and Waterman and Jones (1990) a different approach is taken. The algorithm matches short words of set length and degree of mismatch. The words can be matched within a fixed amount of position offset and total score is maximized where a score is given to each matching word.

3. Secondary structure

When RNA is transcribed from the DNA template, it is single-stranded. That is, RNA does not possess a matching or self-complementary strand to pair with it. The single-stranded molecule can fold back on itself and when regions of the molecule are complementary they can become double-stranded or helical. The pairing rules for the sequences are analogous to those for DNA except that T becomes U (uracil) in the RNA alphabet: A pairs with U and G pairs with C. In addition, frequently G is thought to pair with U. Biologists call the two-dimensional self-pairing *secondary structure*. Without reference to an actual RNA sequence it is an interesting problem to enumerate the distinct secondary structures that are possible under various restrictions suggested by biology. The number of structures for a sequence of length n satisfies a recursion related to the Catalan numbers. There is even a vector recursion that is "Catalan-like".

Next, a definition of secondary structure is given. Label n points on the x-axis: $1, 2, \ldots, n$. The points correspond to the nucleotide sequence of the RNA. Choose a subset of $2j$ points, $0 \leqslant 2j < n$. The $2j$ points are arranged into j disjoint pairs and the pairs are connected by arcs, subject to the following conditions:

1. Adjacent points are never connected by an arc.
2. Any two points connected by an arc must be separated by at least m points.
3. Arcs cannot intersect.

Condition 2 comes from restrictions on the bending of the sugar–phosphate backbone. In RNA $m = 3$ or 4 is realistic. Condition 3 comes from eliminating structures with "knotted" loops. There are a few examples in biology where condition 3 is violated and no combinatorics has yet been done for those cases.

In the next section, computer prediction of secondary structures for RNA sequences is briefly discussed. As was the case for sequence alignment, the associated dynamic programming algorithms are closely related to enumeration of the configurations.

3.1. *Prediction of secondary structure*

Several attempts were made on the secondary structure "problem" before dynamic programming was first proposed. The basic problem is to find the minimum free-energy structure where negative free energy is assigned to the base pairs and positive energy is assigned to end loops, unpaired bases in helical regions, and so on. The energy rules are not too well understood. The subject is reviewed in Zuker and Sankoff (1984) and here a very simple version of the problem is solved: find the secondary structures that have the maximum number of base pairs.

Theorem 3.1. *Let* $x = x_1x_2 \cdots x_n$ *be a sequence over* $\{A, C, G, U\}$, $1 \leqslant m$, *and* $p : \{A, C, G, U\} \times \{A, C, G, U\} \to \{0, 1\}$. *Define* $F(i, j)$ = *maximum number of base pairs of all secondary structures over* $x_i \cdots x_j$, *where a pair can be formed if and only if* $p(\cdot, \cdot) = 1$. *Set* $F(i, j) = 0$ *whenever* $j \leqslant m + i$. *Then*

$$F(i,j) = \max\{F(i,j-1), [F(i,k-1) + F(k+1,j-1) + 1]p(x_k, x_j);\\ 1 + k + m \leqslant j\}.$$

Proof. The proof of the recursion is based on the observation that either x_j is unpaired or it is paired with a base x_k. To satisfy the constraints, $m \leqslant j - k - 1$. The boundary conditions simply reflect the fact that no pairs can form unless the constraint is satisfied. □

The recursion can be performed in $O(n^2)$ time and space. Unfortunately the structures predicted by this algorithm usually do not correspond to those known to exist in nature and more complicated algorithms must be employed. A very useful algorithm has been devised, again based on dynamic programming, that takes time $O(n^3)$ and $O(n^2)$ space (Zuker and Sankoff 1984). This method employs a shortcut and until recently no polynomial-time solution was known for the general problem. In Watermand and Smith (1986) a general solution was given that takes time $O(n^4)$ and space $O(n^3)$. Since sequences of interest are often 5000 long and range up to 20000, there is a need for more work in this area. See Galil and Giancarlo (1989).

3.2. *Counting secondary structures*

Let $S_n(m) = S_n$ be the number of secondary structures possible for a string or sequence of length n. For this discussion the structures need only satisfy the conditions stated above – no sequence-specific pairing is considered. The results are taken from Stein and Waterman (1978) and Howell et al. (1980).

Theorem 3.2. *For* $1 \leqslant m$, $S_0 = S_1 = \cdots = S_{m-1} = 0$ *and* $S_m = 1$ *are boundary values. Then*

$$S_{m+j} = S_{m+j-1} + S_{m+j-2} + \ldots + S_{j-1} + \sum_{0<i\leqslant m+j-2} S_i S_{m+j-2-i}$$

Proof. The proof is similar to the proof of the algorithm given for Theorem 3.1. Consider adding the base $m+j$. If base $m+j$ does not pair we have S_{m+j-1} structures. Otherwise the base pairs with a base with subscript $i+1$ from 1 to $m+j-2$ and the number of structures is the product of the possibilities from the strings $1\cdots i$ and $i+2\cdots m+j-1$. The boundary conditions give the recursion in the above form. □

If we set $m = 0$ and consider the above recursion,

$$S_n(0) = S_n = S_{n-1} + \sum_{0\leqslant j\leqslant n-2} S_j S_{n-2-j}$$

where $S_n(0) = 1$. This recursion generates the Motzkin numbers (Sloan 1973) and they have an explicit solution: $S_n(0) = \sum_{j\geqslant 0} c_{j+1}\binom{n}{2j}$. This formula is a consequence of Theorem 3.3 below, where we define the Catalan numbers c_{j+1} by

$$c_{j+1} = (j+1)^{-1}\binom{2j}{j}.$$

Next define the convolved Fibonacci numbers $f_n(r,k)$ by

$$(1 - x - x^2 - \cdots - {}^{r})^{-k} = \sum_{n\geqslant 0} f_n(r,k)x^n.$$

Th

$$f_n(m+1, 2j+1)x^n.$$

aring the generating function, using the re-
ons. Next asymptotics for S_n are presented.
rem of Bender (1974).

Theorem 3.4. *Define $F(r,s) = r^2s^2 - (1 - r - r^2 - \cdots - r^{m+1})s + r^m$. Let $r > 0, s > S_0$ be the unique real solutions of the system $F(r,s) = 0, F_y(r,s) = 0$. Then*

$$S_n \sim (rF_x(r,s)/(2\pi F_{yy}(r,s)))^{1/2} n^{-3/2} r^{-n}.$$

The following special cases hold:

1. $S_n(0) \sim \sqrt{3/(4\pi)}n^{-3/2}3^n$.
2. $S_n(1) \sim \sqrt{(15+7\sqrt{5})/(8\pi)}n^{-3/2}((3+\sqrt{5})/2)^n$.
3. $S_n(2) \sim \sqrt{(1+\sqrt{2})/\pi}n^{-3/2}(1+\sqrt{2})^n$.

The behavior of $S_n(m)$ is governed by $r(m)^{-n}$ and $r(m)$ can only be numerically determined for $m \geqslant 3$. Still, it can be shown that $r(m)$ is monotonically increasing and $r(m) \to 1/2$ as $m \to \infty$.

Next a closer examination of $S_n(m) = S_n$ is made. Set

$$\boldsymbol{R}^n = (r_0^n, r_1^n, r_2^n, \ldots)$$

where $r_0^n = 1$ and r_i^n is the number of secondary structures with i base pairs for a sequence of length n. Also define $\boldsymbol{a} * \boldsymbol{b} = \boldsymbol{c}$ by

$$c_k = \sum_{0 \leqslant i \leqslant k} a_i b_{k-i},$$

and let $\zeta(a_0, a_1, \ldots) = (0, a_0, a_1, \ldots)$.

Theorem 3.5. *Set $\boldsymbol{R}_0 = \boldsymbol{R}_1 = \boldsymbol{R}_2 = (1,0,0,\ldots)$. Then for $n \geqslant 2$,*

$$\boldsymbol{R}_{n+1} = \boldsymbol{R}_n + \sum_{1 \leqslant j \leqslant n-1} \zeta[\boldsymbol{R}_{j-1} * \boldsymbol{R}_{n-j}].$$

Proof. There can be no pairs for $n \leqslant 2$, so the boundary conditions hold. Next, the number of structures with $i + 1$ pairs for a sequence of length $n + 1$ is derived. If base $n + 1$ is unpaired, s_{i+1}^n structures exist with $i + 1$ pairs. If instead base $n + 1$ is paired with base j, then to have $i + 1$ pairs i additional pairs are needed. If k pairs exist for bases 1 to $j - 1$, then $i - k$ pairs must exist for bases $j + 1$ to n. □

4. Maps of DNA

For many years maps of the relative location of genes on chromosomes have been constructed by a technique known as linkage analysis. Before 1980 it was required that the genes have observable mutations available as genetic markers. Since fruit flies have many visibly distinguishable mutants, the relative locations of many of the corresponding genes have been mapped. Bacteria and yeast have also been extensively mapped. In 1980 it was realized that measurable changes in the DNA itself can be used for genetic mapping (Botstein et al. 1980). The changes are in the lengths of restriction fragments and the resulting mapping techniques have been very important in localizing genes associated with major diseases.

Site-specific restriction enzymes were discovered in 1970 in bacteria (Nathans and Smith 1975). These enzymes cut double-stranded DNA at the locations of short specific patterns, usually from four to six letters in length. The restriction enzyme HhaI cuts at $GCGC$ while EcoRI cuts at $GAATTC$. Mutations at a single letter of DNA can cause the appearance or disappearance of a restriction site. It is easy to see that insertions and deletions of a segment of DNA can also cause variation in the restriction sites. The fragment lengths then become the genetic markers for linkage analysis. This is a quite active research area and there is currently some mathematical activity in devising efficient algorithms (Lander and Botstein 1986).

In linkage analysis, map distance might not relate linearly to physical distance or number of bases. Soon after the discovery of site-specific restriction enzymes biologists learned to construct another type of map known as a restriction map. In restriction maps all the enzyme sites are approximately located on the DNA. Such maps usually cover a few thousand bases of DNA but much longer stretches of DNA have been mapped (Isono et al. 1987). This section will discuss in some detail the difficulties in restriction map construction. First, restriction maps are related to interval graphs.

4.1. Maps as interval graphs

Interval graph theory began with Benzer's study of the structure of genes in bacteria (Benzer 1959). Benzer was able to obtain data on the overlap between pairs of fragments of DNA from a gene. He was successful in arranging the overlap data in a way that implied the linear nature of the gene. Soon after this, Fulkerson and Gross (Golumbic 1980) studied interval graphs and incidence matrices; that study is closely related to Benzer's analysis. Today the linear nature of the gene is well established but interval graphs also arise in connection with restriction maps (Waterman and Griggs 1986).

Representing an interval of DNA as a line segment, the biologist indicates the location of restriction sites along the line segment. Circular DNA does occur in nature but this discussion is restricted to the linear maps of two restriction enzymes. Next the two restriction enzymes are designated by A and B. Figure 1a gives an example of a two-enzyme A/B map while the single-enzyme maps appear in fig. 1b. Biologists are able to measure the lengths but not the order of the intervals between sites, so they are labeled arbitrarily. The intervals are called restriction fragments and form the nodes of our graphs.

Label the ith fragment from enzyme A (B) by A_i (B_i). Define the incidence matrices $I(A,B)$, $I(A,A/B)$, and $I(B,A/B)$ where, for example, $I(A,B)_{i,j} = 1$ if $A_i \cap B_j \neq \emptyset$ and 0 otherwise. It is sometimes experimentally feasible to determine $I(A,A/B)$ and $I(B,A/B)$. It is then a simple matter to compute $I(A,B)$.

Proposition 4.1. *If I^{T} is the transpose of I, then*

$$I(A,B) = I(A,A/B)I^{\mathrm{T}}(B,A/B).$$

Proof. The result follows from the observation that the (i,j)th element of the matrix product is equal to the number of A/B intervals in both A_i and B_j. □

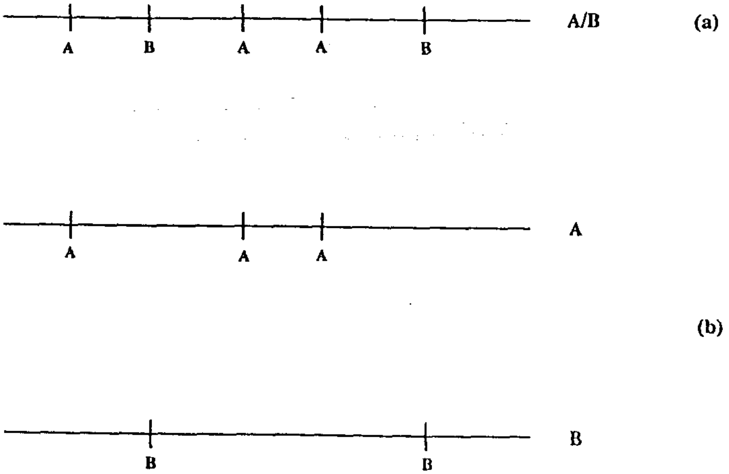


Figure 1. The two enzyme A/B map (a) and the single enzyme (A and B) maps (b).

Constructing a restriction map from $I(A, B)$ is equivalent to finding an interval representation of a bipartite graph $G(A, B)$ defined in a natural way. The vertex set $V(A, B)$ is the union of the set of A intervals with the set of B intervals; the edge set $E(A, B)$ consists of those sets $\{A_i, B_j\}$ where $A_i \cap B_j \neq \emptyset$. If we delete the endpoints of the fragments from the line segment, we obtain an open-interval representation of $G(A, B)$. Restriction maps can be characterized by known results on interval graphs (Golumbic 1980).

Theorem 4.2. *The following are equivalent:*

1. *$G(A, B)$ is a bipartite graph constructed from a restriction map.*
2. *$G(A, B)$ is a bipartite interval graph with no isolated edges.*
3. *$I(A, B)$ can be transformed by row and column permutations into a staircase form with each row or column having 1's in precisely one of the steps.*

With the identification of $G(A, B)$ as an interval graph, it is routine to adapt a general algorithm of Booth and Leuker to recognize G as an interval graph and to give its representation in linear time (Booth and Leuker 1976, Waterman and Griggs 1986). As is the case in many problems in biology, the overlap data is usually given with errors. Then the problem of finding the "interval graph" becomes much harder.

4.2. Constructing maps

It is experimentally possible to apply restriction enzymes singly or in combination, and to estimate the lengths of the resulting fragments of DNA. The problem is to construct the map of location of the enzyme sites along the DNA from this fragment-length data. The results are from Goldstein and Waterman (1987).

4.2.1. *Simulated annealing*

Here we consider the simplest problem of interest that involves linear DNA, two restriction enzymes, and no measurement error. We will refer to this problem as the double-digest problem or problem DDP. A restriction enzyme cuts a piece of DNA of length L at all occurrences of a short specific pattern and the lengths of the resulting fragments are recorded. In the double-digest problem we have as data the list of fragment lengths when each enzyme is used singly, say,

$$\boldsymbol{A} = \{a_i \colon 1 \leqslant i \leqslant n \text{ from the first digest}\},$$
$$\boldsymbol{B} = \{b_i \colon 1 \leqslant i \leqslant m \text{ from the second digest}\},$$

as well as a list of double-digest fragment lengths when the restriction enzymes are used in combination and the DNA is cut at all occurrences specific to both patterns, say

$$\boldsymbol{C} = \{c_i \colon 1 \leqslant i \leqslant n_{1,2}\};$$

only length information is obtained. In general $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ will be multisets; that is, there may be values of fragment lengths that occur more than once. We adopt the convention that the sets $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ are ordered; that is, $a_i \leqslant a_j$ for $i \leqslant j$, and similarly for the sets $\boldsymbol{B}$ and $\boldsymbol{C}$. Of course

$$\sum_{1\leqslant i\leqslant n} a_i = \sum_{1\leqslant i\leqslant m} b_i = \sum_{1\leqslant i\leqslant n_{1,2}} c_i = L,$$

since we are assuming that fragment lengths are measured in number of bases with no errors.

Given the above data, the problem is to find orderings for the sets $\boldsymbol{A}$ and $\boldsymbol{B}$ such that the double-digest implied by these orderings is, in a sense made precise below, $\boldsymbol{C}$. This is a mathematical statement of a problem originally solved by exhaustive search.

The double-digest problem can be stated more precisely as follows. For permutations $\sigma \in (12\cdots n)$, $\mu \in (12\cdots m)$, call (σ, μ) a configuration. By ordering $\boldsymbol{A}$ and $\boldsymbol{B}$ according to σ and μ, respectively, the set of locations of cut sites is obtained:

$$S = \left\{ s \colon s = \sum_{1\leqslant j\leqslant r} a_{\sigma(j)} \text{ or } s = \sum_{1\leqslant j\leqslant t} b_{\mu(j)};\ 0 \leqslant r \leqslant n,\ 0 \leqslant t \leqslant m \right\}.$$

The set S is not allowed repetitions; that is, S is not a multiset. Now label the elements of S such that

$$S = \{s_j \colon 0 \leqslant j \leqslant n_{1,2}\} \quad \text{with } s_i \leqslant s_j \text{ for } i \leqslant j.$$

The double-digest implied by the configuration (σ, μ) can be defined by

$$C(\sigma, \mu) = \{c_i(\sigma, \mu)\colon\ c_i(\sigma, \mu) = s_j - s_{j-1} \text{ for some } 1 \leqslant j \leqslant n_{1,2}\},$$

where it is assumed as usual that the set is ordered in the index i. The problem then is to find a configuration (σ, μ) such that $\boldsymbol{C} = C(\sigma, \mu)$. As discussed below, this problem lies in the class of NP-complete problems conjectured to have no polynomial-time solution.

In order to implement a simulated annealing algorithm, an energy function and a neighborhood structure are required. The energy function is a chi-square-like function

$$f(\sigma, \mu) = \sum_{1 \leqslant i \leqslant n_{1,2}} (c_i(\sigma, \mu) - c_i)^2 / c_i.$$

Note that if all measurements are free of error then f attains its global minimum value of zero for at least one choice (σ, μ). Following Goldstein and Waterman (1987), we define the set of neighbors of a configuration (σ, μ) by

$$N(\sigma, \mu) = \{(\tau, \mu)\colon\ \tau \in N(\sigma)\} \cup \{(\sigma, \nu)\colon\ \nu \in N(\mu)\},$$

where $N(\rho)$ are the neighbors used in studies of the travelling salesman problem (Bonomi and Lutton 1984).

With these ingredients, the algorithm was tested on exact, known data from the bacteriophage lambda with restriction enzymes BamHI and EcoRI, yielding a problem size of $|A|! \times |B|! = 6!6! = 518\,400$. See Daniels et al. (1983) for the complete sequence and map information about lambda. Temperature was not lowered at the rate $c/\log(n)$ as suggested by the theorem in Geman and Geman (1984), but for reasons of practicality was instead lowered exponentially. On three separate trials using various annealing schedules the solution was located after 29 702, 6895, and 3670 iterations from random initial configurations.

4.2.2. Multiplicity of solutions

In many instances, the solution to the double-digest problem is not unique. Consider, for example,

$$\begin{aligned} \boldsymbol{A} &= \{1, 3, 3, 12\}, \\ \boldsymbol{B} &= \{1, 2, 3, 3, 4, 6, \}, \\ \boldsymbol{C} &= \{1, 1, 1, 1, 2, 2, 2, 3, 6\}. \end{aligned}$$

This problem, of size $4!6!/2!2! = 4320$, admits 208 distinct solutions. That is, there are 208 distinct orders which produce $\boldsymbol{C}$. We now demonstrate that this phenomenon is far from isolated.

Below, we use the Kingman subadditive ergodic theorem to prove that the number of solutions to the double-digest problem increases exponentially as a function of length under the probability model stated below.

For reference, a version of the subadditive ergodic theorem is given here (Kingman 1973). For s, t non-negative integers with $0 \leqslant s \leqslant t$, let $X_{s,t}$ be a collection of random variables which satisfy

1. whenever $s < t < u$, $X_{s,u} \leqslant X_{s,t} + X_{t,u}$;
2. the joint distribution of $\{X_{s,t}\}$ is the same as that of $\{X_{s+1,t+1}\}$;
3. the expectation $g_t = E[X_{0,t}]$ exists and satisfies $g_t \geqslant -Kt$ for some constant K and all $t > 1$.

Then the finite $\lim_{t\to\infty} X_{0,t}/t = \lambda$ exists with probability one and in the mean.

Theorem 4.3. *Assume the sites for two restriction enzymes are independently distributed with cut probabilities p_1, p_2 respectively and $p_i \in (0,1)$. Let $Y_{s,t}$ be the number of solutions between the sth and the tth coincident sites. Then there is a constant $\lambda > 0$ such that*

$$\lim_{t\to\infty} \frac{\log(Y_{0,t})}{t} = \lambda.$$

Proof. Let a coincidence be defined to be the event that a site is cut by both restriction enzymes; such an event occurs at each site independently with probability $p_1 p_2 > 0$, and at site 0 by definition. On the sites $1,2,3,\ldots$, there will be an infinite number of such events. For $s, u = 0,1,2,\ldots$, $0 \leqslant s \leqslant u$ we may consider the double-digest problem for only that segment located between the sth and uth coincidences. Let $Y_{s,u}$ denote the number of solutions to the double-digest problem for this segment.

Suppose $s < t < u$. A solution for the segment between the sth and tth coincidences and a solution for the segment between the tth and uth coincidences can be combined to yield a solution for the segment between the sth and uth coincidences. Thus

$$Y_{s,u} \geqslant Y_{s,t} Y_{t,u}.$$

We note that the inequality may be strict as $Y_{s,u}$ counts solutions given by orderings where fragments initially between, say, the sth and tth coincidences now appear in the solution between the tth and uth coincidences. Letting

$$X_{s,t} = -\log Y_{s,t},$$

we have $s \leqslant t \leqslant u$ implies $X_{s,u} \leqslant X_{s,t} + X_{t,u}$.

Additional technical details can be established to complete the proof of the theorem. □

4.2.3. Computational complexity

Given the definition of a restriction map as permutations of the various digest fragments, it is no surprise that the double-digest problem is NP-complete. See Garey and Johnson (1979) for definitions.

Theorem 4.4. *The double-digest problem is NP-complete.*

Proof. It is clear that the DDP described above is in the class NP, as a nondeterministic algorithm need only guess a configuration (σ, μ) and check in polynomial time if $C(\sigma, \mu) = C$. The number of steps to check this is in fact linear. To show that DDP is NP-complete, the partition problem is transformed to DDP.

In the partition problem, known to be NP-complete (Garey and Johnson 1979), a finite set Q, say $|Q| = n$ is given along with a positive integer $s(q)$ for each $q \in Q$, and we wish to determine whether there exists a subset $Q' \subset Q$ such that

$$\sum_{q \in Q'} s(q) = \sum_{q \in Q - Q'} s(q).$$

If $\sum_{q \in Q} s(q) = J$ is not divisible by two, there can be no such subset Q'. Otherwise, input to problem DDP the data

$$\begin{aligned} \boldsymbol{A} &= \{s(a_k) : 1 \leqslant k \leqslant n\}, \\ \boldsymbol{B} &= \{J/2, J/2\}, \\ \boldsymbol{C} &= \{s(q) : q \in Q\}. \end{aligned}$$

It is clear that any solution to problem DDP with this data yields a solution to the partition problem through the order of the implied digest $\boldsymbol{C}$. Therefore DDP is NP-complete. □

Biologists have routinely been solving DDPs for inexact length measurements. They of course are generally unaware of the results presented here, and search for the length and enzymes that allow a solution. To contribute usefully to this field the challenge is to find algorithms that will extend their capabilities.

References

Bender, E.A.

[1974] Asymptotic methods in enumeration, *SIAM Rev.* **16**, 485–515.

Benzer, S.

[1959] On the topology of genetic fine structure, *Proc. Nat. Acad. Sci. U.S.A.* **45**, 1607–1620.

Bonomi, E., and J.L. Lutton

[1984] The N-city travelling salesman problem: Statistical mechanics and the Metropolis algorithm, *SIAM Rev.* 26, 551–568.

Booth, K.S., and G.S. Leuker

[1976] Testing for the consecutive ones property, interval graphs, and graph planarity using PO algorithms, *J. Comput. Syst. Sci.* **13**, 335–379.

Botstein, D., R.L. White, M. Skolnick and R. Davis

[1980] Construction of a genetic linkage map in man using restriction fragment length polymorphisms, *Amer. J. Human Genet.* **32**, 314–331.

Carrillo, H., and D. Lipman

[1988] The multiple sequence alignment problem in biology, *SIAM J. Appl. Math.* **48**, 1073–1082.

Daniels, D., J. Schroeder, W. Szybalski, F. Sanger, A. Coulson, G. Hong, D. Hill, G. Peterson and F. Blattner

[1983] Complete annotated lambda sequence, in: *Lambda II*, eds. R.W. Hedrix, J.W. Roberts and F.W. Weisberg (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).

Galil, Z., and R. Giancarlo

[1989] Speeding up dynamic programming with applications to molecular biology, *Theor. Comput. Sci.* **64**, 107–118.

Garey, M.R., and D.S. Johnson

[1979] *Computers and Intractability: A Guide to the Theory of NP-Completeness* (W.H. Freeman, San Francisco, CA).

Geman, S., and D. Geman

[1984] Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images, *IEEE Trans. Pattern Anal. Mach. Intell.* **6**, 721–741.

Goldstein, L., and M.S. Waterman

[1987] Mapping DNA by stochastic relaxation, *Adv. in Appl. Math.* **8**, 194–207.

Golumbic, M.C.

[1980] *Algorithmic Graph Theory and Perfect Graphs* (Academic Press, New York).

Greene, C.

[1988] Posets of shuffles, *J. Combin. Theory A* **47**, 191–206.

Griggs, J.R., P.J. Hanlon and M.S. Waterman

[1986] Sequence alignments with matched sections, *SIAM J. Algebraic Discrete Methods* **7**, 604–608.

Griggs, J.R., P.J. Hanlon, A.M. Odlyzko and M.S. Waterman

[1990] On the number of alignments of k sequences, *Graphs Combin.* **6**, 133–146.

Howell, J.A., T.F. Smith and M.S. Waterman

[1980] Computation of generating functions for biological molecules, *SIAM J. Appl. Math.* **39**, 119–133.

Isono, K., Y. Kohara and K. Akiyama

[1987] The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library, *Cell* **50**, 495–508.

Kingman, J.F.C.

[1973] Subadditive ergodic theory, *Ann. Probab.* **1**, 883–909.

Kruskal, J.B., and D. Sankoff

[1983] *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison* (Addison-Wesley, Reading, MA).

Lander, E., and D. Botstein

[1986] Strategies for studying heterogeneous genetic traits in humans by using a linkage map of restriction fragment length polymorphisms, *Proc. Nat. Acad. Sci. U.S.A.* **83**, 73–53.

Laquer, H.T.

[1981] Asymptotic limits for a two-dimensional recursion, *Studia Appl. Math.* **64**, 271–277.

Miller, W., and E.W. Myers

[1988] Sequence comparison with concave weighting functions, *Bull. Math. Biol.* **50**, 97–120.

Nathans, D., and H.O. Smith

[1975] Restriction endonucleases in the analysis and restructuring of DNA molecules, *Ann. Rev. Biochem.* **44**, 273–293.

Sloan, N.J.A.

[1973] *A Handbook of Integer Sequences* (Academic Press, New York).

Stein, P.R., and M.S. Waterman

[1978] On some new sequences generalizing the Catalan and Motzkin numbers, *Discrete Math.* **26**, 261–272.

Taylor, W.R.

[1987] Multiple sequence alignment by a pairwise algorithm, *Comput. Appl. Biosci.* **3**, 81–87.

Vingron, M., and P. Argos

[1989] A fast sensitive multiple sequence alignment algorithm, *Comput. Appl. Biosci.* **5**, 115–121.

Waterman, M.S.

[1984] General methods of sequence comparison, *Bull. Math. Biol.* **46**, 473–500.

[1986] Multiple sequence alignment by consensus, *Nucleic Acids Res.* **14**, 9095–9102.

[1989] *Mathematical Methods for DNA Sequences*, ed. M.S. Waterman (CRC Press, Boca Raton, FL).

Waterman, M.S., and J.R. Griggs
[1986] Interval graphs and maps of DNA, *Bull. Math. Biol.* **48**, 189–195.

Waterman, M.S., and R. Jones
[1990] Consensus methods for DNA and protein sequence alignments, in: *Methods in Enzymology*, Vol. 183, ed. R. Doolittle (Academic Press, New York).

Waterman, M.S., and M.D. Perlwitz
[1984] Line geometries for sequence comparisons, *Bull. Math. Biol.* 46, 567–577.

Waterman, M.S., and T.F. Smith
[1986] Rapid dynamic programming algorithms for RNA secondary structure, *Adv. in Appl. Math.* 7, 455–464.

Zuker, M., and D. Sankoff
[1984] RNA secondary structures and their prediction, *Bull. Math. Biol.* 46, 591–622.

CHAPTER 40

Combinatorics in Computer Science

L. LOVÁSZ

Department of Computer Science, Yale University, New Haven, CT 06520, USA

D.B. SHMOYS and É. TARDOS

School of Operations Research and Industrial Engineering, Upson Hall, Cornell University, Ithaca, NY 14853, USA

Contents

HANDBOOK OF COMBINATORICS
Edited by R. Graham, M. Grötschel and L. Lovász

It is not surprising that computer science is perhaps the most important field of applications of combinatorial ideas. Modern computers operate in a discrete fashion both in time and space, and much of classical mathematics must be "discretized" before it can be implemented on computers as, for example, in the case of numerical analysis. The connection between combinatorics and computer science might be even stronger than suggested by this observation; each field has profited from the other. Combinatorics was the first field of mathematics where the ideas and concepts of computer science, in particular complexity theory, had a profound impact. This framework for much of combinatorics has been surveyed in chapter 29. In this chapter, we illustrate what computer science profits from combinatorics: we have collected a number of examples, all of them rather important in computer science, where methods and results from classical discrete mathematics play a crucial role. Since many of these examples rely on concepts from theoretical computer science that have been discussed in chapter 29, the reader is encouraged to refer to that chapter for background material.

1. Communication complexity

There are many situations where the amount of communication between two processors jointly solving a certain task is the real bottleneck; examples range from communication between the earth and a rocket approaching Jupiter to communication between different parts of a computer chip. We shall see that communication complexity also plays an important role in theoretical studies, and in particular, in the complexity theory of circuits. Other examples of such indirect applications of communication complexity include bounds on the depth of decision trees (Hajnal et al. 1988) and pseudorandom number generation (Babai et al. 1989). Communication complexity is a much simpler and cleaner model than computational or circuit complexity, and it illustrates notions from complexity theory like non-determinism and randomization in a particularly simple way. Our interest in this field also stems from the many ways in which it relates to combinatorial problems and methods. This section gives just a glimpse of this theory; see Lovász (1990) for a broader survey.

Suppose that there are two players, Alice and Bob, who want to evaluate a function $f(x,y)$ in two variables; for simplicity, we assume that the value of f is a single bit. Alice knows the value of the variable x, Bob knows the value of the variable y, and they can communicate with each other. Local computation is free, but communication is costly. What is the minimum number of bits that they have to communicate?

We can describe the problem by a matrix as follows. Let $a_1,\ldots,a_N$ be the possible inputs of Alice and $b_1,\ldots,b_M$, the possible inputs of Bob. Note that since local computation is free, we need not worry about how these are encoded. Let $c_{ij}=f(a_i,b_j)$. The 0–1 matrix $C=(c_{ij})_{i=1}^{N}{}_{j=1}^{M}$ determines the communication problem. Both players know the matrix C; Alice also knows the index i of a row, Bob knows the index j of a column, and they want to determine the entry c_{ij}.

To solve their task for a particular matrix C, Alice and Bob, before learning their inputs i and j, agree in advance on a *protocol*, which is the communication analogue to the fundamental notion of an algorithm in computational complexity theory. Informally, a *communication protocol* is a set of rules specifying the order and "meaning" of the messages sent. The protocol prescribes each action for Alice and Bob: who is to send the first bit; depending on the input of that processor, what this bit should be; depending on this bit, who is to send the second bit, and depending on the input of that processor and on the first bit sent, what this second bit should be, and so on. The protocol terminates when one processor knows the output bit and the other one knows this about the first one. The *complexity* of a protocol is the number of bits communicated in the worst case.

The *trivial protocol* is that Alice tells her input to Bob. We shall see that sometimes there is no better protocol than this trivial one. This protocol takes $\lceil \log_2 N \rceil$ bits; if $M < N$ then the reverse trivial protocol is clearly better. (For the remainder of this chapter, we shall use log to denote $\log_2$.)

A protocol has the following combinatorial description in terms of the matrix. First, it determines who sends the first bit; say, Alice does. This bit is determined by the input of Alice; in other words, the protocol partitions the rows of the matrix C into two classes, and the first bit of Alice tells Bob which of the two classes contains her row. From now on, the game is limited to the submatrix C_1 formed by the rows in this class. Next, the protocol describes a partition of either the rows or columns of C_1 into two classes, depending on who is to send the second bit, and the second bit itself specifies which of these two classes contains the line (row or column) of the sender. This limits the game to a submatrix C_2, and so it continues.

If the game ends after k bits then the remaining submatrix C_k is the union of an all-1 submatrix and an all-0 submatrix. We shall call this an *almost-homogeneous matrix*. If, for example, Alice knows the answer, then her row in C_k must be all-0 or all-1, and since Bob knows for sure that Alice knows the answer, this must be true for every row of C_k.

We can therefore characterize the communication complexity by the following combinatorial problem: given a 0–1 matrix C, in how many rounds can we partition it into almost-homogeneous submatrices, if in each round we can split each of the current submatrices into two (either horizontally or vertically). We shall denote this number by $\kappa(C)$.

If $\mathrm{rk}(C)$ denotes the rank of C, then the following inequality, observed by Mehlhorn and Schmidt (1982), relates the communication complexity to the rank of the matrix.

Lemma 1.1. *Over any field,* $\log \mathrm{rk}(C) \leqslant \kappa(C) \leqslant \mathrm{rk}(C)$.

(The lower bound is tightest if we use the real field, whereas the upper bound might be tightened by considering a finite field.)

In particular, we obtain that if C has full row rank then the trivial protocol is optimal. This corollary applies directly to a number of communication problems, of which we mention three. Suppose that Alice knows a subset X of an n-element set S and Bob knows a subset Y of S. The *equality problem* is to decide if the two subsets are equal; the *disjointness problem* is to decide if the two subsets are disjoint; the *binary inner product problem* is to decide if the intersection of the two sets has odd cardinality. In the first two cases, it is trivial to see that the corresponding $2^n \times 2^n$ matrices have full rank; in the third case, the rank of the matrix is $2^n - 1$. Hence the trivial protocols are optimal. (It is interesting to remark that in the third case, the rank over GF(2), which might seem more natural to use in this problem, gives a very poor bound here: the GF(2) rank of C is only n.)

The lower bound in the lemma is often sharp; on the other hand, no communication problem is known for which $\kappa(C)$ is anywhere near the bound $\mathrm{rk}(C)$. In particular, it is open whether $\kappa(C)$ can be bounded by any polynomial of $\log \mathrm{rk}(C)$. Spieker and Raz (1994) constructed an example where $\kappa(C)$ is not linear in $\log \mathrm{rk}(C)$, in fact, $\kappa(C) > \log \mathrm{rk}(C) \cdot \log\log\log \mathrm{rk}(C)$.

To point out that the situation is not always trivial, consider again the equality problem. Recall that if N denotes the number of inputs, then the trivial protocol takes $\lceil \log N \rceil$ bits, and this is optimal.

In contrast, Freivalds (1979) obtained the very nice result that if Alice and Bob tolerate errors occurring with probability 2^{-100}, then the situation changes drastically. Consider the following protocol, which can be viewed as an extension of a "parity check". Treat the inputs as two natural numbers x and y, $0 \leqslant x, y \leqslant N - 1$. Alice selects a random prime $p \leqslant (\log N)^2$, computes the remainder x' of x modulo p, and then sends the pair (x', p) to Bob. Bob then computes the remainder y' of y modulo p, and compares it with x'. If they are distinct, he concludes that $x \neq y$; if they are the same, he concludes that $x = y$.

If the two numbers x and y are equal then, of course, so are x' and y' and so the protocol reaches the right conclusion. If x and y are different, then it could happen that $x' = y'$ and so the protocol reaches the wrong conclusion. This happens if and only if p is a divisor of $x - y$. Since $|x - y| < N$, it follows that $x - y$ has fewer than $\log N$ different prime divisors. On the other hand, Alice had about $(\log N)^2 / 2 \log\log N$ primes from which to choose, and so the probability that she chose one of the divisors of $x - y$ tends to 0. For N sufficiently large, the error will be smaller than 2^{-100}. Clearly, this protocol uses at most $4 \log\log N$ bits.

Randomization need not lead to such dramatic improvements for all problems. We have seen that the binary inner product problem and the disjointness problem behave quite similarly to the equality problem from the point of view of deterministic communication complexity: the corresponding matrices have essentially full rank and hence the trivial protocols are optimal. But, unlike for the equality problem, randomization is of little help here: Chor and Goldreich (1988) proved that the randomized communication complexity of computing the binary inner product problem is $\Omega(n)$. Improving results of Babai et al. (1986), Kalyanasundaram and Schnittger (1987) showed that the randomized communication complexity of the disjointness problem is $\Omega(n)$. The proofs of these facts combine the work of Yao (1983) on randomized communication complexity with rather involved combinatorial considerations.

Just as in computational complexity theory, non-determinism plays a crucial role in communication complexity theory. Non-deterministic protocols were introduced by Lipton and Sedgewick (1981). Perhaps the best way to view a non-deterministic protocol is as a scheme by which a third party, knowing both inputs, can convince Alice and Bob what the value of f is. Again, we want to minimize the number of bits such a *certificate* contains for the worst inputs. For example, if in the disjointness problem the two subsets are not disjoint, announcing a common element convinces both players that the answer is "no". This certificate takes only $\lceil \log n \rceil$ bits, which is much smaller than the number of bits Alice and Bob would need to find the answer, which is n, as we have seen. On the other hand, if the sets are disjoint, then no such simple non-deterministic scheme exists. We shall distinguish between the non-deterministic protocol for the "0" and for the "1". In both cases, there is always the trivial protocol that announces the input of Alice.

To give a formal and combinatorial description of non-deterministic protocols, consider a non-deterministic protocol for "1", a particular certificate p, and those entries of C (i.e., inputs for Alice and Bob) for which this certificate "works". If the proof scheme is correct, these must be all 1's; from the fact that Alice and Bob have to verify the certificate independently, we also see that these 1's must form a submatrix. Thus, a non-deterministic protocol corresponds to a covering of C by all-1 submatrices. Conversely, every such covering gives a non-deterministic protocol: one can simply use the name of the submatrix containing the given entry as a certificate. The number of bits needed is the logarithm of the number of different certificates used: the non-deterministic communication complexity $\kappa_1(C)$ of a matrix C is the least natural number t such that the 1's in C can be covered by at most 2^t all-1 submatrices. One can analogously define $\kappa_0(C)$, the non-deterministic communication complexity of certifying a 0.

Note that the all-1 submatrices in the covering need not be disjoint. Therefore, there is no immediate relation between $\mathrm{rk}(C)$ and $\kappa_1(C)$. It is easy to formulate the non-deterministic communication complexity of a matrix as a set-covering problem: consider the hypergraph whose vertices are the 1's in C and whose edges are the all-1 submatrices; $\kappa_1(C)$ is the logarithm of the minimum number of edges covering all vertices. An immediate lower bound on $\kappa_1(C)$ follows from a simple counting argument: if C has a 1's but each all-1 submatrix of C has at most b entries, then trivially, $\kappa_1(C) \geqslant \log a - \log b$. We can also consider the natural dual problem: if α is the maximum number of 1's in C such that no two occur in the same all-1 submatrix, then $\kappa_1(C) \geqslant \log \alpha$.

Returning to the three problems on two sets mentioned above, we see that the 1's in the $2^n \times 2^n$ identity matrix cannot be covered by fewer that 2^n all-1 submatrices; hence the non-deterministic communication complexity of equality is n. Similarly, the non-deterministic communication complexity of set disjointness is n, since $\alpha = 2^n$. For the binary inner product problem, it is easy to see by elementary linear algebra that an all-1 submatrix has at most 2^{n-1} entries, and that exactly $2^{n-1}(2^n - 1)$ entries are 1's. This gives that at least $2^n - 1$ all-1 submatrices are needed to cover the 1's, and so $\kappa_1(C) = n$. For this matrix, a similar argument also shows that $\kappa_0(C) = n$.

We have derived the following lower bound on the communication complexity of a matrix:

$$\max\{\log \mathrm{rk}(C), \kappa_0(C), \kappa_1(C)\} \leqslant \kappa(C).$$

The identity matrix shows that κ_1 might be a very weak bound: $\kappa(C)$ can be exponentially large compared with $\kappa_1(C)$. Interchanging the roles of 1 and 0, we obtain that $\kappa_0(C)$ might also be very far from $\kappa(C)$. No such example is known for $\log \mathrm{rk}(C)$, but it is likely that the situation is similar.

However, it is a surprising fact that the product of any two of these three lower bounds is an upper bound on the communication complexity. The first part of the following theorem is due to Aho et al. (1983b), and the other two, to Lovász and Saks (1993).

Theorem 1.2. *For every matrix C,*

(a) $\kappa(C) \leqslant (\kappa_0(C)+1)(\kappa_1(C)+1)$;

(b) $\kappa(C) \leqslant 1 + \lfloor \log \mathrm{rk}(C) \rfloor (\kappa_0(C)+2)$;

(c) $\kappa(C) \leqslant 1 + \lfloor \log(\mathrm{rk}(C)+1) \rfloor (\kappa_1(C)+2)$.

We shall sketch a result that is stronger than each of (a), (b) and (c). Let $\rho_1(C)$ denote the size of the largest square submatrix of C such that, after a suitable reordering of the rows and columns, each diagonal entry is 1 but each entry above the diagonal is 0. It is clear that $\rho_1(C) \leqslant \mathrm{rk}(C)$ and $\rho_1(C) \leqslant 2^{\kappa_1(C)}$. If we define $\rho_0(C)$ analogously, then $\rho_0(C) \leqslant \mathrm{rk}(C)+1$. By using these inequalities, we can obtain all three parts of Theorem 1.2 from the following result.

Theorem 1.3. $\kappa(C) \leqslant 1 + \lfloor \log \rho_1(C) \rfloor \big(\kappa_0(C)+1\big)$.

Proof. We use induction on $\rho_1(C)$. If $\rho_1(C) = 1$ then trivially $\kappa(C) \leqslant 1$. Assume that $\rho_1(C) > 1$ and let $k = \kappa_0(C)$; then the 0 elements of C can be covered by all-0 submatrices $C_1, \ldots, C_l$ where $l \leqslant 2^k$. Let A_i denote the submatrix formed by those rows of C that meet C_i, and let B_i denote the analogous submatrix formed from columns of C. Observe that

$$\rho_1(A_i) + \rho_1(B_i) \leqslant \rho_1(C), \qquad i = 1, \ldots, l; \tag{1.1}$$

this will play a crucial role in the following protocol.

First, Alice looks at her row to see if it intersects any submatrix C_i with $\rho_1(A_i) \leqslant \rho_1(C)/2$. If so, she sends a "1" and the name i of such a submatrix. They have thereby reduced the problem to the matrix A_i. If not, she sends a "0".

When Bob receives this "0", he looks at his column to see if it intersects any submatrix C_i with $\rho_1(B_i) \leqslant \rho_1(C)/2$. If so, he sends a "1" and the name i of such a submatrix. They have thereby reduced the problem to the matrix B_i. If not, he sends a "0".

If both Alice and Bob failed to find an appropriate submatrix, then by (1.1), the intersection of their lines cannot belong to any C_i, and so it must be a "1". They have found the answer. □

Theorem 1.2(a) has an interesting interpretation. Define a *communication problem* as a class $\mathcal{H}$ of 0–1 matrices; for simplicity, assume that they are square matrices. The communication complexity of any $N \times N$ matrix is at most $\log N$. We say that $\mathcal{H}$ is in $\mathcal{P}_{\text{comm}}$ if it can be solved substantially better: if there exists a constant $c > 0$ such that $\kappa(C) \leqslant (\log\log N)^c$ for each $N \times N$ matrix $C \in \mathcal{H}$. Similarly, we say that $\mathcal{H}$ is in $\mathcal{NP}_{\text{comm}}$ if there exists a constant $c > 0$ such that for each $C \in \mathcal{H}$, $\kappa_1(C) \leqslant (\log\log N)^c$, and define co-$\mathcal{NP}_{\text{comm}}$ analogously based on $\kappa_0(C)$. Just as for the analogous computational complexity classes, we have the trivial containment

$$\mathcal{P}_{\text{comm}} \subseteq \mathcal{NP}_{\text{comm}} \cap \text{co-}\mathcal{NP}_{\text{comm}}.$$

However, for the communication complexity classes we also have the following, rather interesting facts:

$$\mathcal{P}_{\text{comm}} \neq \mathcal{NP}_{\text{comm}},$$

$$\mathcal{NP}_{\text{comm}} \neq \text{co–}\mathcal{NP}_{\text{comm}},$$

which follow from the equality problem, but

$$\mathcal{P}_{\text{comm}} = \mathcal{NP}_{\text{comm}} \cap \text{co–}\mathcal{NP}_{\text{comm}},$$

by Theorem 1.2(a). This idea was developed by Babai et al. (1986), who defined and studied communication analogues of many other well-known complexity classes such as $\#\mathcal{P}$, $\mathcal{PSPACE}$ and $\mathcal{BPP}$.

Our previous examples were trivial from the point of view of communication: the trivial protocols are optimal. This is quite atypical. Let us define a *round* in the protocol as a maximal period during which one player sends bits to the other. The trivial protocol consists of one round. Let $\kappa^k(C)$ denote the minimum number of bits needed by a communication protocol with at most k rounds. Halstenberg and Reischuk (1988) proved that for every $k \geqslant 1$, there exist arbitrarily large matrices C such that $\kappa^k(C)$ is exponentially larger than $\kappa^{k+1}(C)$.

We consider two combinatorial examples to illustrate that protocols can be significantly more efficient than the trivial ones. Yannakakis (1988) considered the following problem: Alice and Bob are given a graph G on n nodes, Alice is given a stable set A and Bob is given a clique B; they must decide whether these sets intersect. The corresponding matrix C has rank n but size exponential in n, and so the trivial protocol takes $\Omega(n)$ bits. (Recall that we always focus on the worst-case complexity.) On th r hand, the non-deterministic communication complexity
n, since the name of a common node of A and B is a
2, $\kappa(C) \leqslant (2 + \log n)^2$. (The number of rounds in the pro-
t known whether the deterministic complexity, or even
plexity of disjointness, is smaller, such as $\mathrm{O}(\log n)$. It is
he latter question is equivalent to the following purely
there a constant $c > 0$ so that in each graph G on n
ich that each pair of disjoint sets U, V, where U is a
is separated by one of these cuts.

In the *subtree disjointness problem*, there is a tree T known to both players. Alice gets a subtree T_A and Bob gets a subtree T_B, and their task is to decide whether T_A and T_B are node-disjoint. It can be shown that the corresponding matrix C has rank $2n$ but has exponential size. The non-deterministic complexity of non-disjointness is $\log n$, and hence by Theorem 1.2, $\kappa(C) \leqslant (2+\log n)^2$. In this case, one can do better by using the following simple protocol: Alice sends any node x of her tree to Bob. If y is the node in Bob's tree that is closest to x, Bob responds by sending y to Alice. Then Alice checks if $y \in T_A$: if so, then clearly the subtrees are not disjoint; if not, the subtrees are disjoint. This protocol uses $2\log n$ bits. Lovász and Saks (1993) showed that it can be modified to use only $\log n + \log\log n$ bits.

An interesting and rather general class of communication problems, for which good bounds on the complexity can be obtained by non-trivial combinatorial means, was formulated by Hajnal et al. (1988). Let $(\mathscr{L}, \wedge, \vee)$ be a finite lattice. Assume that Alice is given an element $a \in \mathscr{L}$ and Bob is given an element $b \in \mathscr{L}$, and their task is to decide whether $a \wedge b = 0$, the minimal element of the lattice.

This problem generalizes both the disjointness problem (where $\mathscr{L}$ is a Boolean algebra), and the subtree disjointness problem (where $\mathscr{L}$ is the lattice of subtrees of a tree). A third special case worth mentioning is the following *spanning subgraph problem*: Alice is given a graph G_A and Bob is given a graph G_B on the same set of nodes V; they wish to decide whether $G_A \cup G_B$ is connected. This case relies on the lattice of partitions of V, but "upside down" so that the indiscrete partition is 0. Alice can compute the partition a of V into the connected components of G_A, Bob can compute the partition b of V into the connected components of G_B, and then they decide whether $a \wedge b = 0$.

For a given lattice $\mathscr{L}$, let C be the matrix associated with the corresponding problem: its rows and columns are indexed by the elements of $\mathscr{L}$, and $c_{ij} = 1$ if and only if $i \wedge j = 0$. To find the rank of this matrix, we give the following factorization of it, using the Möbius function μ of $\mathscr{L}$ (see chapter 21 for the definition and some basic properties). Let $Z = (z_{ij})$ be the zeta-matrix of the lattice, i.e., let $z_{ij} = 1$ if and only if $i \leqslant j$ $(i, j \in \mathscr{L})$. Let $D = (d_{ij})$ denote the diagonal matrix defined by $d_{ii} = \mu(0, i)$. Then it is easy to verify the following identity, found by Wilf (1968) and Lindström (1969) (see chapter 31):

$$C = Z^{\mathrm{T}} D Z. \tag{1.2}$$

Since Z is trivially non-singular, this implies that

$$\mathrm{rk}(C) = \mathrm{rk}(D) = |\{i\colon \mu(0, i) \neq 0\}|. \tag{1.3}$$

This gives a lower bound on the communication complexity of our problem, but how good is this bound? One case when this bound is tight occurs when $\mu(0, i) \neq 0$ for all i. We obtain a lower bound of $\lceil \log |\mathscr{L}| \rceil$, which is also the upper bound achieved by the trivial protocol; that is, the trivial protocol is optimal. By Corollary 3.10 of chapter 21, this case occurs when $\mathscr{L}$ is a geometric lattice or a geometric lattice "upside down". In particular, we see that for the spanning subgraph problem, the trivial protocol is optimal.

It turns out that the lower bound given by $\log \mathrm{rk}(C)$ is not too far from the truth for any lattice.

Theorem 1.4. *For every lattice $\mathscr{L}$,*

$$\log \mathrm{rk}(C) \leqslant \kappa(C) \leqslant (1 + \log \mathrm{rk}(C))^2.$$

Proof (*upper bound*). Observe that a non-deterministic certificate of non-disjointness of two elements $a, b \in \mathscr{L}$ can be provided by exhibiting an atom of the lattice below both a and b. Hence the logarithm of the number of atoms is an upper bound on $\kappa_0(C)$. Since for every atom i, $\mu(0, i) = -1$, it follows from the identity (1.3) that the number of atoms is at most $\mathrm{rk}(C)$. Hence, the upper bound follows directly from Theorem 1.2. □

2. Circuit complexity

One promising approach to proving lower bounds on the computational complexity of a problem focuses on the Boolean circuit model of computation, and recent results in this area are possibly the deepest applications of combinatorial methods to computer science thus far. The best way to view a circuit is not as an abstract electronic device; instead, view it as the bit-operational skeleton of any computational procedure. This way, it is not hard to see that this model is equivalent to other models such as the Turing machine or RAM, and that the number of functional elements, or gates, in a circuit is equivalent to the time taken by an algorithm in those models. (More precisely, a RAM algorithm, for example, is equivalent to a family of Boolean circuits, one for each input length.) As a result, the extremely difficult task of proving lower bounds on the computational complexity of a given problem can be posed in a way much more suited to combinatorial methods. Many people believe that this is the direction of research that may eventually lead to the solution of famous problems, such as $\mathscr{P}$ vs. $\mathscr{NP}$. Unfortunately, the handful of results obtained at this point are rather difficult, and yet quite far from this objective.

Let us recall some definitions from chapter 29. A *Boolean circuit* is an acyclic directed graph; nodes of indegree 0 are called *input gates*, and are labelled with the input variables $x_1, \ldots, x_n$; nodes with outdegree 0 are called *output gates*; every node with indegree $r > 0$ is called a *functional gate*, and is labelled with a Boolean function in r variables, corresponding to the predecessors of the node. For our purposes, it suffices to allow only the logical negation, conjunction, and disjunction as functions. The number of gates in a circuit is called its *size*; the *circuit complexity* of a problem is the size of the smallest circuit that computes this Boolean function. The outdegree and indegree of a node are referred to as its *fan-out* and *fan-in*, respectively.

Another important parameter of a circuit is its *depth*, the maximum length of a path from an input gate to an output gate. A circuit can also be viewed as parallel algorithm, and then the depth corresponds to the (parallel) time that the algorithm takes. Note that every Boolean function can be computed by a Boolean circuit of depth 2: this is easily derived from its conjunctive normal form. Of course, the number of gates, which is essentially the number of terms in this normal form, is typically exponential.

A circuit is *monotone* if only conjunctions and disjunctions are allowed as functional gates. Note that every monotone increasing Boolean function can be computed by a monotone Boolean circuit.

The predominant approach to proving circuit complexity lower bounds is to restrict the class of allowed circuits. Two kinds of restrictions have proved sufficiently strong, and yet reasonably interesting, to allow the derivation of superpolynomial lower bounds on the number of gates: monotonicity and bounding the depth. Two main methods seem to emerge: the random restriction method and the approximation method. Both methods have applications for both kinds of restricted problems.

The first superpolynomial lower bound in a restricted model of computation concerned constant-depth circuits. Note that for this class of circuits to make sense, we must allow that the gates have arbitrarily large fan-out and fan-in. Furst et al. (1981) and Ajtai (1983) proved independently that every constant-depth circuit computing the parity function has superpolynomial size; the *parity* function maps $(x_1, \dots, x_n) \mapsto x_1 + \cdots + x_n$, where here, and throughout this section, addition is the mod 2 sum. Yao (1985) established a truly exponential lower bound by extending the techniques of Furst et al. Hastad (1989) has further strengthened the bound and greatly simplified the proof. All of these proofs are based on probabilistic combinatorial arguments; the proof of the following theorem can be found in chapter 33.

Theorem 2.1. *If C is a circuit with n input bits and depth d that computes the parity function, then C has at least* $2^{(1/10)n^{1/(d-1)}}$ *gates.*

Razborov (1987) gave an exponential lower bound on the size of constant-depth circuits that compute another simple Boolean function, the so-called *majority function*, i.e., the function

$$f(x_1, \dots, x_n) = \begin{cases} 1, & \text{if at least } \lceil n/2 \rceil \text{ of the } x_i \text{ are } 1, \\ 0, & \text{otherwise.} \end{cases}$$

In fact, he proved a stronger result by allowing circuits that may have *parity gates*, in addition to the usual AND, OR and NOT gates, where a parity gate computes the parity of the number of 1's in its input. The proof uses the *approximation method*, which was first used by Razborov (1985a) in his pathbreaking paper on monotone circuits. The later application of the method is perhaps the cleanest, and we are able to reproduce the full proof.

Theorem 2.2. *If C is a circuit of depth d, with* AND, OR, NOT *and parity gates that computes the majority function of n input bits, then C has at least* $2^{n^{(1/2d)}}/10\sqrt{n}$ *gates.*

Proof. Consider a circuit that computes the majority function. We can assume without loss of generality that the circuit uses only parity and OR gates, since these can be used to simulate both AND and NOT gates within constant depth. The idea of the proof is to introduce "approximations" of the gates used during the computation. Using the approximate gates, instead of the real gates, one computes an approximation of the majority function. The quality of the approximation will be measured in terms of the number of inputs on which the modified circuit differs from the original. The main point of the approximation is to keep the computed function "simple" in some sense. We will show that every "simple" function, and in particular the approximation we compute, differs from the majority function on a significant fraction of the inputs. Since the approximation of each gate has a limited effect on the function computed, we can conclude that many approximations had to occur.

Each Boolean function can be expressed as a polynomial over the two-element field GF(2). The measure of simplicity of a Boolean function f for this proof is the degree of the polynomial representing the function or for short, the *degree of the function*.

In fact, the approximation technique is applied not to the majority function, but to a closely related function, the *k-threshold function* f_k. This function is 1 when at least k of the inputs are 1. It is easy to see that if there is a circuit of size s that computes the majority function of $2n-1$ elements in depth d, then, for each k, $1 \leqslant k \leqslant n$, there is a circuit of depth d and size at most s that computes the k-threshold function on n elements. Therefore, any exponential lower bound for f_k implies a similar bound for the majority function. We shall consider $k = \lceil (n+h+1)/2 \rceil$ for an appropriate h.

First we show that any function of low degree has to differ from the k-threshold function on a significant fraction of the inputs.

Lemma 2.3. *Let $n/2 \leqslant k \leqslant n$. Every polynomial with n variables of degree $h = 2k-n-1$ differs from the k-threshold function on at least $\binom{n}{k}$ inputs.*

Proof. Let g be a polynomial of degree h and let $\mathscr{B}$ denote the set of vectors where it differs from f_k. Let $\mathscr{A}$ denote the set of all 0–1 vectors of length n containing exactly k 1's.

For each Boolean function f, consider the summation function $\hat{f}(x) = \sum_{y \leqslant x} f(y)$. It is trivial to see that the summation function of the monomial $x_{i_1} \cdots x_{i_r}$ is 1 for the incidence vector of the set $\{i_1, \ldots, i_r\}$ and 0 on all other vectors. Hence it follows that f has degree at most h if and only if $\hat{f}$ vanishes on all vectors with more than h 1's. In contrast to this, the summation function of the k-threshold f_k is 0 on all vectors with fewer than k 1's, but 1 on all vectors with exactly k 1's.

Consider the matrix $M = (m_{ab})$ whose rows are indexed by the members of $\mathscr{A}$, whose columns are indexed by the members of $\mathscr{B}$, and

$$m_{ab} = \begin{cases} 1, & \text{if } a \geqslant b, \\ 0, & \text{otherwise.} \end{cases}$$

We want to show that the columns of this matrix generate the whole linear space. This will imply that $|\mathcal{B}| \geqslant |\mathcal{A}| = \binom{n}{k}$.

Let $a_1, a_2 \in \mathcal{A}$ and let $a_1 \wedge a_2$ denote their coordinatewise minimum. Then we have, by the definition of $\mathcal{B}$,

$$\sum_{\substack{b\leqslant a_1\\ b\in\mathcal{B}}} m_{a_2b} = \sum_{\substack{b\leqslant a_1\wedge a_2\\ b\in\mathcal{B}}} 1 = \sum_{u\leqslant a_1\wedge a_2} (f_k(u)+g(u)) = \sum_{u\leqslant a_1\wedge a_2} f_k(u) + \sum_{u\leqslant a_1\wedge a_2} g(u).$$

The second term of this last expression is 0, since $a_1 \wedge a_2$ contains at least $h+1$ 1's. The first term is also 0 except if $a_1 = a_2$. The columns of M therefore generate the unit vector corresponding to the coordinate a_1, and so they generate the whole space. □

If p_1 and p_2 are polynomials representing two functions, then $p_1 + p_2$ is the polynomial corresponding to the parity of the two functions. The polynomial p_1p_2 corresponds to their AND, which makes it easy to see that $(p_1+1)(p_2+1)+1$ corresponds to their OR. Note that the inputs have degree 1, i.e., they are very simple. Since the degree is not increased by computing the sum, the parity gates do not have to be approximated. On the other hand, unbounded fan-in OR gates can greatly increase the degree of the computed functions. We will approximate the OR gates so that the approximated function will have fairly low degree. The following lemma will serve as the basis for the approximation.

Lemma 2.4. *Let $g_1, \ldots, g_m$ be Boolean functions of degree at most h. If $r \geqslant 1$ and $f = \bigvee_{i=1}^{m} g_i$, then there is a function f' of degree at most rh that differs from f on at most 2^{n-r} inputs.*

Proof. Randomly select r subsets $I_j \subseteq \{1, \ldots, m\}$ $(1 \leqslant j \leqslant r)$, where each i is independently included in I_j with probability $\frac{1}{2}$. Let f_j be the sum of the g_i with $i \in I_j$, and consider $f' = \bigvee_{j=1}^{r} f_j$. We claim that the probability that f' satisfies the requirements of the lemma is non-zero. It is clear that the degree of the polynomial for f' is at most rh. Furthermore, consider an input α; we claim that the probability that $f'(\alpha) \neq f(\alpha)$ is at most 2^{-r}. To see this, consider two cases. If $g_i(\alpha) = 0$ for every i, then both $f(\alpha) = 0$ and $f'(\alpha) = 0$. On the other hand, if there exists an index i for which $g_i(\alpha) = 1$, then $f(\alpha) = 1$ and for each j, $f_j(\alpha) = 0$ independently with probability at most $\frac{1}{2}$. Therefore, $f'(\alpha) = 0$ with probability at most 2^{-r}, and the expected number of inputs on which $f' \neq f$ is at most 2^{n-r}. Hence for at least one particular choice of the sets I_j, the polynomial f' differs from f on at most 2^{n-r} inputs. □

To finish the proof, assume that there is a circuit of size s and depth d to compute the k-threshold function for inputs of size n. Now apply Lemma 2.4 with $r = \lfloor n^{1/(2d)} \rfloor$ to approximate the OR gates in this circuit. The functions computed by the gates at the ith level will be approximated by polynomials of degree at most r^i. Therefore, each resulting approximation p_k of the k-threshold function will have degree at most r^d. Lemma 2.3 implies that for $k = \lceil (n + r^d + 1)/2 \rceil$, the polynomial p_k differs from the k-threshold function on at least $\binom{n}{k}$ inputs. This shows that $s2^{n-r} \geqslant \binom{n}{k}$. From this, routine calculations yield that

$$s \geqslant \binom{n}{k} 2^{r-n} > \frac{2^r}{\sqrt{\pi n}},$$

which establishes the desired exponential lower bound. □

Smolensky (1987) generalized this result to prove that every constant-depth circuit that decides whether the sum of the inputs is 0 modulo p using AND, OR and modulo-q gates has exponential size, where p and q are powers of different primes.

How far can one relax the assumption on bounded depth and still obtain superpolynomial lower bounds? The methods of Yao, Hastad and Razborov can be extended to depth near $\log n/\log\log n$. One cannot hope for much more, since the parity function can in fact be computed by a linear-size circuit of depth $\mathrm{O}(\log n/\log\log n)$.

Perhaps the deepest result on circuit complexity is contained in the groundbreaking paper of Razborov (1985a). He gave a superpolynomial lower bound on the monotone circuit complexity of the clique problem, without any restriction on the depth. Shortly afterwards, Andreev (1985) used similar techniques to obtain an exponential lower bound on a less natural $\mathcal{NP}$-complete problem. Alon and Boppana (1987), by strengthening the combinatorial arguments of Razborov, proved an exponential lower bound on the monotone circuit complexity of the k-clique function.

Theorem 2.5. *If C is a monotone circuit with $\binom{n}{2}$ input bits that decides whether a given graph on n nodes contains a clique with at least s nodes, then the number of gates in C is at least*

$$\tfrac{1}{8}\left(\frac{n}{4s^{3/2}\log n}\right)^{(\sqrt{s}+1)/2}.$$

The proof uses a much more elaborate application of the approximation technique. The main combinatorial tool used is the "sunflower theorem" of Erdős and Rado (see chapter 24).

How different can the monotone and non-monotone circuit complexity of a monotone function be? Pratt (1975) proved that the monotone circuit complexity of Boolean matrix multiplication is $\Theta(n^3)$. This, together with the $\mathrm{O}(n^{\log 7})$ matrix multiplication technique of Strassen (1969), proves that these two notions are distinct. Razborov (1985b), using techniques similar to those used for the clique lower bound, showed that the perfect matching problem, which is in $\mathcal{P}$, has superpolynomial monotone circuit complexity, thereby establishing a superpolynomial gap. Tardos (1988) showed that this could be increased to an exponential separation, by combining the arguments of Razborov (1985a), Alon and Boppana (1987) and results of Grötschel et al. (1981) on the polynomial computability of a graph function ϑ that is closely related to the clique function (see chapter 31).

As remarked above, no methods are known to handle general circuits with depth greater than $\log n$. In the case of monotone circuits with fan-in 2, however, a version of the random restriction method has been successfully applied by Karchmer and Wigderson (1990) to prove a lower bound on the depth proportional to $\log^2 n$. It is clear that a circuit with fan-in 2 and size N must have depth at least $\log N$; hence Theorem 2.5 implies that every monotone circuit with fan-in 2 computing the k-clique function must have depth $n^{1/3-\varepsilon}$. However, the function that Karchmer and Wigderson consider is computable by polynomial-size monotone circuit, and so no non-trivial bound on the depth is implied by considering only its size.

Karchmer and Wigderson considered the *undirected reachability problem*: given a graph G and two nodes s and t, is there an $s-t$ path in G? This problem is clearly in $\mathcal{P}$, and in fact, it can be decided by a polynomial-size monotone circuit that has depth $O(\log^2 n)$. Karchmer and Wigderson proved the following result.

Theorem 2.6. *There exists a constant $c > 0$ such that if C is a monotone circuit with fan-in 2 that solves the undirected reachability problem for a graph on n nodes, its depth is least $c \log^2 n$.*

The proof, which uses a version of the random restriction method, is quite involved and is not given here. We describe, however, the starting point, which is a new characterization of the depth of circuits with fan-in at most 2 in terms of communication complexity, thereby establishing a surprising link with the material of the previous section.

Consider the following game between two players. The game is given by a Boolean function f in n variables. The first player gets $x \in \{0,1\}^n$ such that $f(x) = 0$ and the second player gets $y \in \{0,1\}^n$ such that $f(y) = 1$. The goal of the game is that the two players should agree on a coordinate i such that $x_i \neq y_i$. Let $\kappa(f)$ denote the minimum number of bits that the two players must communicate to agree on such a coordinate. (For example, the first player could tell x to the second player, and then the second player can find an appropriate coordinate to tell to the first player, so $\kappa(f) \leqslant n + \log n$.) Karchmer and Wigderson proved that the minimum depth in which a Boolean function f can be computed with a circuit with fan-in 2 is equal to $\kappa(f)$. A similar characterization can be given for the monotone circuit complexity of monotone Boolean functions.

In the case of the undirected reachability problem, the corresponding game can be phrased as follows: the first player is given an $[s,t]$-path and the second player is given an $[s,t]$-cut, and the goal of the game is to find an edge in the intersection of the path and the cut. Consider the following protocol for this connectivity game: the first player sends the name of the midpoint on the path, and the second player responds by telling on which side of the cut this node lies. This protocol requires $O(\log n)$ rounds, and in each round the first player sends $\log n$ bits and the second player sends 1.

Karchmer and Wigderson prove that even if the second player were allowed to send $O(n^{\varepsilon})$ bits in each round (instead of just 1), the players would still need at least $\Omega(\log n)$ rounds. The claimed lower bound on the monotone circuit depth is a consequence of this fact.

3. Data structures

Imagine a huge science library. It contains a wealth of information, but to make this information useful, catalogues, reference and review volumes, indices (and librarians) are needed. Similarly, information in the memory of a computer is useful only if it is accessible, i.e., it is provided with extra structures that make the storage, retrieval, and modification of this information possible, and in fact easy. This is particularly important when implementing complicated algorithms: the fast storage and retrieval of certain partial results of the computation is often the main bottleneck in speeding up such algorithms. Such auxiliary structures, called *data structures*, are becoming increasingly important as the amount of information stored increases.

The theory of data structures is very broad and we shall restrict ourselves to two examples that illustrate the depth of combinatorial ideas used in this field. For a more thorough treatment, see Aho et al. (1983a), Tarjan (1983) and Gonnet (1984).

3.1. Shortest paths and Fibonacci heaps

Let $G = (V(G), E(G))$ be a graph with n nodes and m edges, and with a specified node s. Let every edge e have a non-negative length $c(e)$. We want to find a shortest path from s to every other node. We have seen in chapter 2 that using Dijkstra's algorithm, this is quite easily done in polynomial ($O(n^2)$) steps. (To be precise, we use the RAM machine model of computation, and a step means an arithmetic operation – addition, multiplication, comparison, storage or retrieval – of numbers whose length is at most a constant multiple of the maximum of $\log n$ and the length of the input parameters.) Let us review this procedure.

We construct a subtree T of G, one node at a time. We shall only be concerned with the nodes of this tree, so we consider T as a set of nodes. Initially, we let $T = \{s\}$. At any given step, for each $v \in T$ we know the length of the shortest path from s to v, i.e., the distance $d(s,v)$. It would be easy to also obtain the edges of the tree; then the unique $[s,v]$-path in this tree realizes this distance.

The essence of Dijkstra's algorithm is to find an edge $uv \in E(G)$ with $u \in T$ and $v \in V(G) \setminus T$ for which $d(s,u) + c(uv)$ is minimal, and then add v to T. As shown in chapter 2, we then have that $d(s,v) = d(s,u) + c(uv)$. The issue is to find this edge economically. At first glance, we have to select the smallest member from a set of size $O(m)$, and we have to repeat this n times, so this rough implementation of Dijkstra's algorithm takes $O(mn)$ steps.

We can easily do better, however, by keeping track of some of the partial results that we have obtained. Let S denote the set of nodes not in T but adjacent to T. For each node $v \in S$, we maintain the value $\ell(v) = \min\{d(s,u) + c(uv)\}$, where the minimum is taken over all edges uv with $u \in T$. For $v \notin S \cup T$, we define $\ell(v) = \infty$ for notational convenience. Clearly, $\ell(v)$ is an upper bound on $d(s,v)$. At the beginning, $\ell(v) = c(sv)$ for each neighbor v of s and $\ell(u) = \infty$ for each other node u. A step then consists of (a) selecting the node $v \in S$ for which $\ell(v)$ is minimum and setting $d(s,v) = \ell(v)$, (b) deleting v from S and adding it to T, (c) adding each neighbor w of v that is in $V(G) \setminus (T \cup S)$ to S, and (d) updating the value $\ell(w)$ for each neighbor w of v that is in S by

$$\ell(w) := \min\left\{\ell(w), d(s,v) + c(vw)\right\}.$$

This way, it takes $O(n)$ steps to select the node $v \in S$ for which $\ell(v)$ is minimum, and so the total number of steps spent on selecting these nodes is $O(n^2)$. There is also the time needed to update the values $\ell(w)$: this is a constant number of steps per node, and we have at most $d(v)$ nodes to consider, where $d(v)$ is the degree of v. Updating therefore takes $O(\sum_v d(v)) = O(m)$ steps, which is dominated by $O(n^2)$; by maintaining the current best path lengths, Dijkstra's algorithm takes $O(n^2)$ steps.

Can we do better? It is natural to assume that we have to take at least m steps, in order to read the data. If m is proportional to n^2, the algorithm just described is best possible. But for most real-life problems, the graph is sparse, i.e., m is much smaller than n^2, and then there is space for improvement. To obtain this improvement, we shall store the set S, together with the values $\ell(v)$, in a clever way.

A first idea is to keep the set S sorted in order of the value of $\ell(v)$. This makes it trivial to select the next node v and delete it from S, but to achieve (c) and (d), we insert new items in the sorted list, which can be done in $O(\log n)$ steps per insertion. However, even this is non-trivial. If we simply store the sorted elements of S in an array, i.e., in consecutive fields, then to insert a new element, we expect to move half of the old elements for each insertion. Another possibility is to store these elements in a *list*: each element is stored along with a *pointer*, which specifies the memory location of the next element in the list. In this data structure, insertion is trivial, but to find the point of insertion, we must traverse the list, which takes a linear number of steps. Advantages of both methods can be combined using a data structure called a *binary search tree*. We shall not discuss these in detail, since we will show how to do better with another kind of data structure, called a heap. Nonetheless, Dijkstra's algorithm with the current best path lengths stored in a binary search tree takes $O(m \log n)$ steps. This may be much better than $O(n^2)$, but there is still room for improvement.

At this point, it is worth while to formulate the requirements of the desired data structure in an axiomatic way. We have some objects, the elements of S, which are to be stored. Each object has a value $\ell(v)$ associated with it, which is called its *key*. Reviewing the algorithm, we see that we must perform the following operations on this collection of data:

(3.1) DELETEMIN. We might want to find the element of S with smallest key and delete it from S (steps (a) and (b)).

(3.2) INSERT. We might want to add new elements to S (step (c)).

(3.3) DECREASEKEY. We might change the key of an element of S; in fact, we only need to decrease it (step (d)).

Observe that (3.1) and (3.2) are performed $O(n)$ times, since every node is added at most once and deleted once, whereas (3.3) is performed $O(m)$ times.

As mentioned above, a heap is a data structure that can handle these operations in logarithmic time. Since DECREASEKEY is performed more often than the other two operations, we can improve the overall running time by decreasing the cost of performing just this operation. Fredman and Tarjan (1987) showed how to do this by using a more sophisticated data structure, called a Fibonacci heap.

A *heap* is a rooted tree defined on the elements of S, with the property that the key of any node is no more than the key of any of its children. In particular, the root is an element with the smallest key. If the tree is a single path, then the heap is a sorted list, but it will be worth while to consider more compact trees.

Before deciding about the shape of the heap, let us discuss how to perform the tasks (3.1)–(3.3). The heap itself can be realized by maintaining a pointer from each non-root node to its parent. Moreover, the children of each node are ordered in an arbitrary way, and each child contains a pointer to the previous child as well as to the next child. Each parent maintains pointers to its first and last child. Each node has five pointers, some of which may be undefined: parent, first child, last child, previous sibling, next sibling. Changes in the heap are made by manipulating these pointers.

The most common way to implement operations (3.1)–(3.3) in a heap is as follows. DECREASEKEY is perhaps the easiest. Assume that we decrease the key of an element p_0. Let $p_0p_1\cdots p_t$ be the path in the tree connecting p_0 to the root. If $\ell(p_0)$ is still at least $\ell(p_1)$ then we still have a heap; otherwise, we interchange the elements p_0 and p_1. Next we compare the key of p_0 with the key of p_2; if $\ell(p_0) \leqslant \ell(p_2)$, we have a heap; otherwise, interchange p_0 and p_2, and so on. After at most t interchanges we end up with a heap. Note that the tree has not changed, only the vertices have been permuted.

INSERT can be reduced to DECREASEKEY: if we want to add a new element w then we can assign to it a temporary key $+\infty$, and make it the child of any preexisting element. Trivially, this produces a heap. We can then decrease the key of w to the right value, and reorder the elements as before.

Finally, DELETEMIN can be performed as follows. Let r be the root element of the heap. Select any leaf p of the tree and replace r by p. This interchange deletes the root, but we do not necessarily have a heap, since the key of p may be larger than the key of one or more of its children. Find the child with smallest key, and interchange that child with p. It is easy to see that the resulting tree again can violate the heap condition only at the node p. If p has larger key than some of its children, then find its child with the smallest key and interchange them; and so on. Eventually, we obtain a heap.

The *height* of a node in a rooted tree is the maximum distance to a leaf; the height of the tree is the height of its root. If the tree has height h then the operations INSERT and DECREASEKEY take $O(h)$ steps; to make these operations efficient, we want the tree as short as possible. But DELETEMIN puts limits on this: it also involves $O(h)$ interchanges, but before each interchange, we must also find the child with smallest key, and this takes roughly d steps for a node with d children. If we use *balanced k-ary trees* in which, with at most one exception, all internal nodes have k children and each leaf is at distance h or $h-1$ from the root, then $h \leqslant \log_k n$ and so the total number of steps is $O(n \log_k n + m \log_k n + nk \log_k n)$. The best choice is $k = 2 + (m/n)$, and this shows that Dijkstra's algorithm with the current best path lengths stored in a k-ary heap takes $O(m \log n / \log(2 + (m/n)))$ steps, which is a slight improvement.

One can use more sophisticated trees with a more sophisticated implementation of the basic operations and with a more sophisticated way to count steps. A rooted tree is called a *Fibonacci tree* if

$$\text{for every node } v \text{ and natural number } k\text{, the number of children of } v \text{ with degree at most } k \text{ is at most } k+1. \tag{3.4}$$

(We use the term degree in the graph-theoretic sense: the degree of a non-root node is one larger than the number of its children.)

The *degree* of the tree is the degree of the root. We want to build heaps on such trees. For technical reasons, it will be convenient to add an artificial element r with key $-\infty$; hence r is always the root. Moreover, for the root we shall impose the following condition, stronger than (3.4):

$$\text{The degrees of the children of } r \text{ are distinct.} \tag{3.5}$$

A *Fibonacci heap* is a heap whose underlying tree is a Fibonacci tree that satisfies condition (3.5). We are going to show that by using Fibonacci heaps, we can implement Dijkstra's algorithm to run in $O(m + n \log n)$ steps, which is best possible for every $m \geqslant n \log n$. For very sparse graphs, this is, in some sense, an optimal implementation of Dijkstra's algorithm, but other algorithms may be better.

Note that the subtree of a Fibonacci tree formed by a node and its descendants is also a Fibonacci tree. If we delete a node and its descendants, then the only node where condition (3.4) could be violated is the grandparent of the deleted node. In particular, if we delete a child of the root and its descendants, we are left with a Fibonacci tree. By applying induction to this observation, we obtain the following lemma, which explains the name Fibonacci tree.

Lemma 3.1. *Let $F_0 = 0$ and $F_1 = 1$. Then the number of nodes in a Fibonacci tree of degree k is at least F_{k+2}, the $(k+2)$nd Fibonacci number.*

It follows that a Fibonacci tree with n nodes has degree $O(\log n)$. More generally, each node has degree $O(\log n)$, which follows by considering the Fibonacci tree formed by this node and its descendants.

Assume now that we have a Fibonacci heap, and we want to perform INSERT, DELETEMIN and DECREASEKEY operations. In each case, we will first produce a heap that satisfies the Fibonacci property at all non-root nodes; we will then give a procedure that restores the stronger property (3.5) at the root.

(a) INSERT. Add the new node x as a child of r.

(b) DELETEMIN. Of course, we do not want to delete the root, but the minimal "real" element. Find the child of the root with the smallest key, delete it, and make its children have the root as their new parent.

(c) DECREASEKEY. Suppose that we want to decrease the key of a node v_1. If v_1 is a child of the root, simply decrease the key. Otherwise, let $v_1 v_2 \cdots v_t r$ be the path connecting v_1 to the root. Delete the edge connecting v_1 to its parent v_2, and let v_1 become a child of the root. The resulting tree satisfies the heap condition even with the decreased key. Moreover, condition (3.4) is satisfied by all non-root nodes except possibly by v_3, the grandparent of v_1. Until condition (3.4) is satisfied at all non-root nodes, fix the violation at v_j by making v_{j-1} a child of the root. After at most t steps, we obtain a tree that satisfies (3.4) for all non-root nodes.

(d) For each of the three operations, we finish by restoring property (3.5). To do this, we look at the degrees of the root's children, and assume that two children u and v have the same degree t. Suppose that the key of u is no more than the key of v; then we delete the edge vr and make v a child of u. We will show that property (3.4) remains valid at every non-root. This is trivial for all nodes except u. To see that it holds for u, consider any $s \geqslant 1$. If $s < t$ then u has the same children with degree at most s as before, and so their number is at most $s + 1$. For $s \geqslant t$, u now has at most s children altogether. (It will be important that a nearly identical argument applies even if a child of v were deleted!(*))

If we find two children of the root with the same degree then we can transform the heap into one where (3.4) is still valid at all non-roots, and the root has lower degree. We repeat this until all children of the root have different degrees; then (3.4) is trivially satisfied at the root.

There are two points to clarify.

- How do we know in (c) how many edges v_iv_{i+1} must be deleted? To check condition (3.4) for each v_i would take too much time. Instead, we will classify each node as either *safe* or *unsafe*. If a node is safe, then deleting one of its children will not violate (3.4) at a non-root node. In contrast, classifying a node as unsafe implies only that we are not sure if such a violation would occur. Thus, we can always classify the root and its children as safe. In particular, each newly inserted node is safe. We shall reclassify a node in only two cases:

(i) If an unsafe node becomes a child of the root (in a DELETEMIN or DECREASEKEY operation), then it is reclassified as safe.

(ii) If a safe node different from the root loses a child (in a DECREASEKEY operation), then it is reclassified as unsafe.

It follows that in the DECREASEKEY operation, we delete all edges of the path $v_1v_2 \cdots v_tr$ up to the first safe node v_j, and make $v_1, \ldots, v_{j-1}$ children of the root. We reclassify v_j as unsafe, and $v_1, \ldots, v_{j-1}$ as safe. Note that in the parenthetical remark (*), we have already indicated that when r gets a new grandchild v in step (d), it is correct to still classify v as safe.

- In performing (d), how do we find those children of the root with the same degree? To sort the degrees would be too time-consuming. We wish to perform (d) in O(d) steps, where d is the degree of the root after performing steps (a)–(c). For each node, we can always store its current degree in an array. We will also maintain an array a, where $a[k]$, $k = 1, \ldots, d$, indicates the name of a child of the root of degree k, if one exists. This array is not changed during steps (a)–(c), but is updated during step (d) instead. It is trivial to update a to reflect the deletion of a child of the root. To update a for the new children of the root added during steps (a)–(c), we consider them one at a time. To update a to reflect the next child u, if u has degree k then we check if $a[k]$ contains the name of a node. If not, we let $a[k] := u$. If it already contains the name of a child v, then we make one of u and v the child of the other, and update $a[k]$ accordingly. This creates a child of degree $k+1$, which must then be checked with $a[k+1]$, and so forth. Since each "collision" of two children with the same degree causes the degree of r to decrease, there are fewer than d collisions overall, and so the new children are added in $O(d)$ time. In fact, if adding t children causes c collisions, the total work of step (d) is proportional to $t + c$.

We must still bound the time needed to perform these operations. It is easy to see from Lemma 3.1 that INSERT and DELETEMIN operations take $O(\log n)$ steps. But a DECREASEKEY operation may take enormous time! For example, if the (Fibonacci) tree is a single path from the root, with all nodes but the root and its child unsafe, then it takes about n steps to decrease the key of the bottom node. Furthermore, if the root has roughly $\log n$ children, then adding just a single child of the root could take roughly $\log n$ steps.

The remedy is a book-keeping trick called *amortization of costs*. Imagine that we maintain the Fibonacci heap as a service. The customer may order any of the INSERT, DELETEMIN and DECREASEKEY operations. For each operation, we ought to charge him the actual cost, say a cent for each step. But suppose that we also require that he pay a deposit of one dollar for each unsafe node that is created, and a deposit of 25 cents for each child of the root that is created. If either the number of unsafe nodes or the number of children of the root decreases, the appropriate deposit is refunded. With this billing system, an INSERT or DELETEMIN operation still costs only $O(\log n)$ cents, but now we can bound the cost (to him) of a DECREASEKEY operation. Let t be the number of nodes to be made children of the root; this is at most one larger than the number of unsafe nodes becoming safe. Since in step (c), at most one node becomes unsafe, the customer then gets a net refund of at least $75t - 200$ cents. The actual cost of step (c) is proportional to t, certainly at most $50t$. The actual cost of step (d) is proportional to $t + c$, certainly at most $25(t + c)$. However, in step (d), the customer also gets a refund of $25c$ cents. The total cost to the customer of steps (c) and (d) is at most 2 dollars: with this billing system, a DECREASEKEY operation costs only a constant amount.

Summarizing, we have shown the following theorem.

Theorem 3.2. *In a Fibonacci heap, performing n INSERT, n DELETEMIN and m DECREASEKEY operations takes* $O(n \log n + m)$ *steps.*

For our original problem we get the following result.

Theorem 3.3. *Dijkstra's algorithm can be implemented, using Fibonacci heaps, in* $O(n \log n + m)$ *steps.*

3.2. Shortest spanning trees and the UNION–FIND problem

Many of the data structures discussed in the previous subsection can also be used in computing a shortest spanning subtree of a graph. In particular, Fibonacci heaps can be used to implement Prim's algorithm to run on a connected graph G with n nodes and m edges in $O(m + n\log n)$ time. However, in some cases, we may already know the sorted order of the lengths of the edges, or can find this sorted order extremely quickly, such as when the lengths are known to be small integers. In these cases, we can obtain an even more efficient algorithm by using Kruskal's algorithm implemented with a data structure with surprising combinatorial complications. For the following discussion, assume that the sorted order of the edge lengths is known in advance.

Kruskal's algorithm is very simple: it takes the edges one-by-one in the given sorted order, and it adds the next edge to a set T if it does not form a circuit with the edges already in T; otherwise, it disposes of the edge. This seems to take m steps, except that we must check whether the new edge forms a circuit with T. Let uv be the edge considered. To search T for a $[u,v]$-path would be too time-consuming; it would lead to an $O(mn)$ implementation of Kruskal's algorithm.

We can do better by maintaining the partition $\{V_1,\ldots,V_\kappa\}$ of $V(G)$ induced by the connected components of the graph $(V(G),T)$. Each iteration amounts to checking whether u and v belong to the same class; if not, we add uv to T. Furthermore, updating the partition is simple: adding uv to T will cause the two classes containing u and v to merge.

To implement Kruskal's algorithm efficiently, we must therefore find a good way to store a partition of $V(G)$ so that the following two operations can be performed efficiently:

(3.9) FIND. Given an element u, return the partition class containing u.

(3.10) UNION. Given two partition classes, replace them by their union.

We assume that each partition class has a name, and for our purposes, it will be convenient to use an arbitrary element of the class to name the partition class. We shall call this element the *boss*. In a UNION operation, we can keep either one of the old bosses as the new boss. Kruskal's algorithm uses $2m$ FIND operations and $n-1$ UNION operations.

A trivial way to implement a UNION–FIND structure is to maintain a pointer for each element, pointing to the boss of its class. A FIND operation is then trivial; it takes only one step. On the other hand, to do a UNION operation we may have to re-direct almost n pointers, which yields an $O(n^2)$ implementation of Kruskal's algorithm. This is unsatisfactory for sparse graphs, and so we must do the UNION operation more economically.

An almost trivial observation already yields a lot: when merging two classes, we re-direct the pointers in the smaller class. We call this rule *selection by size*. To estimate the number of steps needed when using this rule, observe that whenever a pointer is re-directed, the size of the class containing it gets at least doubled. Hence each pointer is redirected at most $\log n$ times. The total number of steps spent on UNION operations is therefore $O(n\log n)$, and we get an $O(m + n\log n)$ implementation of Kruskal's algorithm. For $m = \Omega(n\log n)$, this is optimal.

To be able to do better for really sparse graphs (e.g., with a linear number of edges), we use a more sophisticated way to keep track of the boss. We shall maintain a rooted tree on each class, with the boss as its root. Each UNION operations then takes only constant time: we make one boss the child of the other. But this makes the FIND operation more expensive: we have to walk up in the tree to the root, so it may take as much time as the height of the tree. This suggests that we should keep the trees short. We can use an analogue to the selection by size rule, called *selection by height*: when merging the two trees, if r_1 is the root of greater height, then it is made the parent of r_2, the root of the other tree. This increases the height only if the two trees had the same height.

Of course, it is time-consuming to compute the height of the trees at each UNION operation, but we can maintain the height $h[v]$ for each node v. This is easily updated: it changes only if the union of two trees is performed and the roots have $h[r_1] = h[r_2]$; then 1 is added at the new root. It is easy to verify by induction that the number of elements in a tree with root r is at least $2^{h[r]}$. Hence $h[r] \leqslant \log n$ for every r, and the cost of a FIND operation is $O(\log n)$. This does not yet yield any improvement in the implementation of Kruskal's algorithm.

But we can use another idea, called *path compression*. Suppose that we perform a FIND operation which involves traversing a fairly long path $v_1 \cdots v_k r$. Then we can traverse this path again, and make each v_i the child of the root. This doubles the number of steps, but the tree becomes shorter.

We combine this idea with a variant of selection by height, called *selection by rank*. For each element v, we maintain a number $\rho[v]$ called its *rank*. The rank of each node is initially 0; if two trees with roots r_1 and r_2 are merged, where $\rho[r_1] \geqslant \rho[r_2]$, we make r_2 a child of r_1, and update ρ by setting

$$\rho[v] := \begin{cases} \rho[r_1] + 1, & \text{if } v = r_1 \text{ and } \rho[r_2] = \rho[r_1], \\ \rho[v], & \text{otherwise.} \end{cases}$$

A path compression does not change ρ. The number $\rho[v]$ is no longer the height of v, but it will be an upper bound on the height. Moreover, the number of elements in a tree with root v is still at least $2^{\rho[v]}$. We shall need the following generalization of this fact, which also follows by induction.

Lemma 3.4. *The number of elements v with $\rho[v] = t$ is at most $n/2^t$.*

For each leaf v, $\rho[v] = 0$, ρ is strictly increasing along any path to the root. Note that this guarantees that the height of the tree is at most $\log n$.

Tarjan (1975) showed that using selection by rank with path compression reduces the cost substantially: the average cost of a FIND operation grows only very, very slowly.

Let us recall Ackerman's function from chapter 25. First, we define a series of functions $f_i : \mathbb{N} \to \mathbb{N}$ by the double recurrence

$$f_1(n) = 2n, \qquad f_i(0) = 2 \quad (i \geqslant 2),$$

$$f_{i+1}(n) = f_i(f_{i+1}(n-1)).$$

Thus, $f_2(n) = 2^{n+1}$, $f_3(n)$ is roughly a tower of $n+1$ 2's, and so on. Ackerman's function is the diagonal $A(n) = f_n(n)$. This grows faster than any f_i, and in fact, faster than any primitive recursive function. The *inverse Ackerman function* $\alpha(n)$ is defined by

$$\alpha(n) = \min\{k\colon A(k) \geqslant n\}.$$

This function grows extremely slowly (e.g., slower than $\log\log n$ or $\log^* n$). We shall also need to introduce inverse functions ϕ_i of the f_i, defined by

$$\phi_i(n) = \max\{k\colon f_i(k) \leqslant n+2\}.$$

The definition is made so that for fixed $n \geqslant 3$, $\phi_i(n)$ is strictly decreasing as a function of i until it reaches 0. Tarjan's result, in a slightly weakened form, is as follows.

Theorem 3.5. *In a rooted forest with n elements, using selection by rank and path compression, m FIND operations and at most $n-1$ UNION operations take $O((n+m)\alpha(n))$ steps.*

Proof. A UNION operation takes only a constant number of steps. To analyze the FIND operation, we first develop a few tools. For each non-root node v, let v' denote its parent. We define the *level* of a node v as the least i for which $\phi_i(\rho[v]) = \phi_i(\rho[v'])$; a root is on level 0. (Note that if $\phi_i(\rho[v]) = \phi_i(p[v'])$ then $\phi_{i+1}(\rho[v]) = \phi_{i+1}(\rho[v'])$.) Initially, each node is therefore on level 0. When a node becomes a non-root, it then reaches a positive level. Note that from this step on, $\rho[v]$ does not change. The value $\rho[v']$ can change in only two ways: v may get a new parent (from a path compression) or v has the same parent and yet $\rho[v']$ changes (from a UNION operation). In either case, $\rho[v']$ increases. Hence the level of the node v can only increase with time. On the other hand, the level of a node remains very small: it is bounded by the least i for which $\phi_i(n) = \phi_i(0)$, which is easily seen to be at most $2\alpha(n+3)$, which is $O(\alpha(n))$.

Let $v_1 \cdots v_k r$ be a path that is being compressed, where r is a root. The cost of this compression is proportional to k; using the charging analogy again, let us say at most k dollars. Consider a node v_j $(1 \leqslant j \leqslant k)$ on the path, and let i be its level. If $\phi_{i-1}(\rho[r]) > \phi_{i-1}(\rho[v_{j+1}])$, then charge one dollar to the node. Otherwise, charge one dollar to the customer.

How much do we charge to the customer per path compression? If he is charged for a node v_j at level i, then by the monotonicity of ϕ_{i-1} and ρ, we see that $\phi_{i-1}(\rho[v_{j+1}]) = \phi_{i-1}(\rho[v_{j+2}]) = \cdots = \phi_{i-1}(\rho[r])$. This implies that $v_{j+1}, \ldots, v_k, r$ are at level $i-1$ or lower, i.e., v_j is the last node at level i on the path. So the customer gets charged for at most one node at each level, which is a total charge of $O(\alpha(n))$. For m path compressions, this is a total of $O(m\alpha(n))$.

How much charge is left on the nodes? While a node v stays on level i, the value $\phi_{i-1}(\rho[v'])$ increases whenever v is charged a dollar, and so the total charge it accumulates is bounded by the maximum value $\phi_{i-1}(\rho[v'])$ can attain. Note that

$\phi_i(\rho[v']) = \phi_i(\rho[v])$ by the definition of level i, and so from the definition of ϕ_i we have

$$\begin{aligned}\rho[v'] + 3 &\leqslant f_i\big(\phi_i(\rho[v']) + 1\big) = f_{i-1}\Big(f_i\big(\phi_i(\rho[v'])\big)\Big) = f_{i-1}\Big(f_i\big(\phi_i(\rho[v])\big)\Big)\\ &\leqslant f_{i-1}(\rho[v] + 2),\end{aligned}$$

and hence,

$$\phi_{i-1}(\rho[v']) \leqslant \rho[v] + 1.$$

So the charge accumulated by a node v while at level i is at most $\rho[v] + 1$, and since $\rho[v]$ never decreases, we can use the final value in this bound. Adding this over all nodes, we can use Lemma 3.4 to show that the total contribution of level i to the charges to the nodes is at most

$$\sum_v 1 + \rho[v] \leqslant n + \sum_t t\frac{n}{2^t} < 3n.$$

Since there are $\mathrm{O}\big(\alpha(n)\big)$ levels, the total charge to the nodes is $\mathrm{O}\big(n\alpha(n)\big)$. □

Corollary 3.6. *Kruskal's algorithm, with pre-sorted edges, can be implemented in* $\mathrm{O}\big(m\alpha(n)\big)$ *steps.*

For a discussion of other implementations of Kruskal's algorithm and its relatives, see Tarjan (1983).

4. Cryptography and pseudorandom numbers

Let $f : \{0,1\}^* \to \{0,1\}^*$ be a one-to-one function, and assume that this function can be evaluated in polynomial time. Can the inverse function be evaluated in polynomial time? We do not know the answer, but it is quite likely that the answer is in the negative; there are some suspected examples (and some constructions that have been disproved). It turns out that such *one-way functions*, and their extensions, could be used in various important applications of complexity theory. We examine some of these applications and constructions.

4.1. Cryptography

The basic goal of cryptography is to provide a means for private communication in the presence of adversaries. Until fairly recently, cryptography has been more of an art than a science. No guarantees were given for the security level of the codes, and in fact, all the classical codes were eventually broken. Modern cryptography has suggested ways to use complexity theory for designing schemes with provable levels of security. We will not be able to discuss in detail, or even mention, most of the results here. We intend rather to give a flavor of the problems considered and the results proved. The interested reader is referred to the surveys by Rivest (1990), Goldreich (1988), Goldwasser (1989), or to the special issue of SIAM Journal on Computing (April, 1988) on cryptography.

A traditional solution uses secret-key crypto-systems. Suppose that two parties wish to communicate a message M, a string of 0's and 1's of length n, and they agree in advance on a secret key K, a string of 0's and 1's of the same length as M. They send the encrypted message $C = M \oplus K$ instead of the real message M, where $\oplus$ denotes the componentwise addition over GF(2). This scheme is provably secure, in the sense that, assuming the key K is kept secret, an adversary can learn nothing about the message by intercepting it: every string of length n is equally likely to be the message encoded by the string C. Unfortunately, it is very inconvenient to use this system since before each transmission the two parties must *agree* on a new secret key.

In the paper that founded the area of modern cryptography, Diffie and Hellman (1976) suggested the idea of public-key crypto-systems. They proposed that a transmission might be sufficiently secure if the task of decryption is computationally infeasible, rather than impossible in an information theoretic sense. Such a weaker assumption might make it possible to securely communicate without *agreeing* on a new secret key every time, and also to achieve a variety of tasks previously considered impossible.

We illustrate the idea of security through intractability by a simple example. Assume that you have a bank account that can be accessed electronically by a password. If the password is long enough, and you keep it safely, then this is secure enough. Except that the programmer of the computer may inspect the memory, learn your password, and then access your account. It would seem that there is no protection against this: the computer has to remember the password, and a good hacker can learn anything stored in the computer.

But the computer does not have to know your password! When you open your account you first generate a Hamiltonian circuit H on, say 10000 nodes, and then add edges arbitrarily to obtain a graph G. The computer of the bank will only store the graph G while your password is the code of a Hamiltonian circuit H. When you punch in your password the computer checks whether it is the code of a Hamiltonian circuit in G, and lets you access the account if it is. For you, the Hamiltonian circuit functions like a password in the usual sense.

But what about the programmer? He can learn the graph G; but to access your account, he should specify a Hamiltonian circuit of G, which is a computationally intr table task. This intractability protects your account!

ucial point here is that given a Hamiltonian circuit, it is easy to construct
taining it, but given a graph, it can be quite hard to find a Hamiltonian
nfortunately, it is easy to find a Hamiltonian circuit in most graphs,
ay known to produce graphs for which finding the Hamiltonian
n be difficult.

s motivate the following definition. A *one-way function* is a
$1\}^* \to \{0,1\}^*$ such that f can be computed in polynomial
e algorithm can invert f on even a polynomial fraction
iven a one-way function f, we can choose an $x \in$

$\{0,1\}^*$ as our password and store the value $f(x)$ in the computer. The length of the word chosen is the *security parameter* of the scheme. This is indeed done in practice. (The above scheme with Hamiltonian circuits leads to a somewhat weaker notion: a "one-way relation", but for most applications, we need a one-way function.)

The password scheme above is just a simple example of what can be achieved by a *public-key crypto-system*. This can have any number of participants. The participants agree on an encryption function Φ, a decryption function Ψ, and a security parameter n. Messages to be sent are divided into pieces of length n. The system functions as follows:

- Each participant should randomly choose a public encryption key E and a corresponding secret decryption key D depending on the security parameter n. A directory of the public keys is published.
- There must be a deterministic polynomial-time algorithm that, given a message M of length n and an encryption key E, produces the encrypted message $\Phi(E,M)$.
- Similarly, there must exist a deterministic polynomial-time algorithm that, given a message M of length n and a decryption key D, produces the decrypted message $\Psi(D,M)$.
- It is required that for every message M, $\Psi(D,\Phi(E,M))=M$.

And the crucial security requirement:

- One cannot efficiently compute $\Psi(D,M)$, without knowing the secret key D. More precisely, for every constant c and sufficientlylarge n, the probability that a (randomized) polynomial-time algorithm using the public key E but not the private key D can decrypt a randomly chosen message of length n is less than n^{-c}.

When a user named Bob wants to send a message M to another user named Alice, he looks up Alice's public-key E_{A} in this directory, and sends her the encrypted message $\Phi(E_{\mathrm{A}},M)$. By the assumptions, only Alice can decrypt this message using her private key D_{A}.

The question is: how do we find such encryption and decryption functions? The basic ingredient of such a system is a "trapdoor" function. The encryption function $\Phi(E,\cdot)$ for a fixed participant must be easy to compute but hard to invert, i.e., a one-way function; but in order for a one-way function to be useful in the above scheme, we need a further feature: it has to be a *trapdoor function*, which is a 2-variable function like Φ in the scheme above: $\Phi\colon\{0,1\}^n\times\{0,1\}^n\to\{0,1\}^n$ such that for every $E\in\{0,1\}^n$, $f(E,\cdot)$ is one-to-one and easily computable, but its inverse is difficult to compute, unless one knows the "key" D belonging to E.

Given the present state of complexity theory, there is little hope to prove that any particular function is one-way or trapdoor (or even that one-way functions exist), or that a public-key crypto-system satisfies the last requirement above. Note that the existence of a one-way function would imply that $\mathcal{P}\neq\mathcal{NP}$. Therefore, a more realistic hope would be to prove the security of a crypto-system based on the assumption that $\mathcal{P}\neq\mathcal{NP}$. However, this seems to be quite difficult for two reasons. Complexity theory is mainly concerned with worst-case analysis. For the purpose of cryptography, average-case analysis, and a corresponding notion of completeness (such as the one suggested by Levin 1986) would be more appropriate. Furthermore, one-way functions lead to languages in $\mathcal{NP}$ which have a unique "witness", whereas the nondeterministic algorithms for $\mathcal{NP}$-complete problems have several accepting paths for most instances.

Over the last ten years, several public-key crypto-systems have been suggested and analyzed. Some have been proven secure, based on assumptions about the computational difficulty of a particular problem (e.g., factoring integers), or on the more general assumption that one-way functions exist.

Rivest et al. (1978) were the first to suggest such a scheme. Their scheme is based on the assumed difficulty of factoring integers. Each user of this scheme has to select a number n that is the product of two random primes. Random primes can be selected using a randomized primality testing algorithm, since every $(\log n)$th number is expected to be prime. The public key consists of a pair of integers (n,e) such that e is relative prime to $\phi(n)$, where $\phi(n)$ denotes the number of integers less than n relatively prime to n. A message M is encrypted by $C = M^e \pmod n$. The private key consists of integers (n,d), where $d \cdot e \equiv 1 \pmod{\phi(n)}$, and decryption is done by computing $C^d = M \pmod n$. Given the prime factorization of n, one can find the required private key d in polynomial time, but the task of finding the appropriate d given only n and e is equivalent to factoring.

Rabin (1979) suggested a variation on this scheme in which any algorithm for decryption, rather than any algorithm for finding the decryption key, can be used to factor an integer. The idea is to encrypt a message M by $C = M^2 \pmod n$. (A slight technical problem that has to be overcome before turning this into an encryption scheme is that squaring modulo a product of two primes is not a one-to-one function, but rather is four-to-one.) An algorithm that can extract square roots modulo n can be used to factor: choose a random integer x and let the algorithm find a square root y of the integer $(x^2 \bmod n)$; with probability $\frac{1}{2}$, the greatest common divisor of $x - y$ and n is one of the two primes used to create n.

Unfortunately, the problem of factoring an integer is not known to be $\mathcal{NP}$-hard; it would be conceptually appealing to suggest schemes that are based on $\mathcal{NP}$-complete problems. Merkle and Hellman (1978) and since then several others, have suggested schemes based on the subset sum problem. The public key of such systems is an integer vector $a = (a_1, \ldots, a_n)$. A message M, that is a 0–1 vector of length n, is encrypted by the inner product $C = (a \cdot M)$. One problem with this scheme is that there does not appear to be a private key that can make the task of decryption easier for the intended receiver. Crypto-systems based on this idea have built some additional trapdoor information into the structure of the vector a, and so the decryption problem is based on a restricted variant of the subset sum problem, which is not known to be $\mathcal{NP}$-hard. Furthermore, randomly chosen subset sum problems can be easy to solve. In an innovative paper, Shamir (1982) used Lenstra's integer programming algorithm to break the Merkle–Hellman scheme; that is, he gave a polynomial-time decryption algorithm that does not need the secret key. Since then, several other subset sum-based schemes have been broken by clever use of the LLL basis reduction algorithm (see chapter 19).

The schemes mentioned so far have encrypted messages of length n, for some given security parameter n. An apparent problem with this approach is that even if we prove that no polynomial-time algorithm can decrypt the messages, it still might be possible to deduce some relevant information about the message in polynomial time. To formalize what it means to exclude this, Goldwasser and Micali (1984) suggested the following framework for a randomized encryption procedure and its security. Suppose that there is a function $B:\{0,1\}^n \to \{0,1\}^m$ and a randomized polynomial-time algorithm that, for any image $B(x)$, produces a value y such that $B(y)=B(x)$, and y is randomly distributed among all such values; an m-bit message can be encrypted by running this algorithm. Intuitively, this encoding is secure if, for any two messages M_1 and M_2 and some pair of codes for them, E_1 and E_2, (i.e., $B(E_i)=M_i$, $i=1,2$) it is "impossible" to gain any advantage in guessing which message corresponds to each encryption. More precisely, the randomized encryption based on B is *secure* if, for each pair (M_1,M_2) any randomized polynomial-time algorithm that is used to guess, given (M_1,M_2) and an unknown random ordering of E_1 and E_2, which value comes from which argument, the probability that the algorithm answers correctly is $\frac{1}{2}+O(n^{-c})$ for all but an n^{-c} fraction of the encryptions, for every $c>0$. Note that if the encryption algorithm is a deterministic polynomial-time algorithm, then it is not secure, since one could simply apply the encryption algorithm to M_1 and M_2 to determine the correspondence.

Goldwasser and Micali proposed a way to achieve this level of security by encrypting messages bit by bit. How does one encode a single bit? A natural solution is to append a number of random bits to the one bit of information, and encrypt the resulting longer string. While this was later shown to be an effective approach, Goldwasser and Micali proposed a different randomized bit-encryption scheme based on a function $B:\{0,1\}^n \to \{0,1\}$ and proved its security, based on a number-theoretic assumption, namely on the assumed difficulty of the quadratic residuosity problem.

An integer x is a *quadratic residue* modulo n if $x=y^2 \pmod n$ for some integer y. The *quadratic residuosity problem* is to decide for given integers n and x whether x is a quadratic residue modulo n. If n is a prime then it is easy to recognize if x is a quadratic residue modulo n, e.g., by computing $x^{(n-1)/2}$ modulo n: this is 0 or 1 if and only if x is a quadratic residue.

The quadratic residuosity problem for composite moduli is more difficult. One has to be careful, however, since there is a polynomially checkable necessary condition that could help in certain cases. To formulate this, define the *Legendre symbol* for any prime n by

$$\left(\frac{x}{n}\right)=\begin{cases}1, & \text{if } n \nmid x \text{ and } x \text{ is a quadratic residue mod } n,\\ -1, & \text{if } n \nmid x \text{ and } x \text{ is not a quadratic residue mod } n,\\ 0, & \text{if } n \mid x.\end{cases}$$

We have seen above that it is easy to compute the Legendre symbol (where n is a prime). The *Jacobi symbol* is a generalization of the Legendre symbol to composite n, but not in the obvious way: if $n = p_1 \cdots p_k$ (where the p_i are not necessarily distinct primes), then

$$\left(\frac{x}{n}\right) = \prod_{i=1}^{k} \left(\frac{x}{p_i}\right).$$

It cannot be seen from this definition, but the Jacobi symbol can be computed in polynomial time for any n.

Now if n and x are coprime integers then $\left(\frac{x}{n}\right) = 1$ is a necessary condition for x to be a quadratic residue modulo n, but it is not sufficient: if n is a product of two primes, then exactly half of the residue classes x with $\left(\frac{x}{n}\right) = 1$ are quadratic residues. The Goldwasser–Micali encryption scheme is based on the assumption that there is no efficient way to obtain further information on the quadratic residuosity of $x \pmod{n}$. It works as follows: a public key consists of an integer n that is the product of two large primes, and a quadratic non-residue y with $\left(\frac{y}{n}\right) = 1$. The bit 0 is encrypted by a random quadratic residue $r^2 \pmod{n}$, the bit 1 is encrypted by a random quadratic non-residue of the form $r^2 y \pmod{n}$. The task of distinguishing encryptions of 0's from encryptions of 1's is exactly the quadratic residuosity problem. Decryption is easy for the intended receiver, who knows the prime factorization of n.

Yao (1982) has extended this result by proving that a secure randomized bit-encryption scheme exists if a one-way function exists; in fact, from every one-way function one can extract a "secure bit". The following simple construction was given by Goldreich and Levin (1989). Let $f: \{0,1\}^* \to \{0,1\}^*$ be a length-preserving one-way function, where a function is *length-preserving* if it maps n-bit strings into n-bit strings, for all n. Define a Boolean function by $B(x) = f^{-1}(x') \cdot x''$, where, if $n = |x|$, then x' and x'' are the first and last $\lfloor n/2 \rfloor$ bits of x, and "$\cdot$" denotes inner product. (If we also want to decode this bit, we take a trapdoor function for f.)

Diffie and Hellman noticed that under a further assumption, a public-key crypto-system can also be used to solve the *signature problem*, where each participant wants a way to electronically sign its messages so that no one else can forge it, in the sense that each recipient can verify that the message must have been signed by the claimed sender. The assumption, which is not very restrictive, is that $\Psi(D, \cdot)$ is one-to-one, i.e., $\Phi(E, \Psi(D, M)) = M$ for each message M. In this case, the system can be used for signatures in the following way. When Bob sends a message M to Alice he can use his private decryption key D_B to append the "signature" $\Psi(D_B, M)$ after the message M. Given such a signature, Alice can use Bob's public key E_B to convince herself that the message came from Bob, exactly as she received it. The Rivest, Shamir and Adleman scheme has this additional property, and therefore can be used for signatures as well.

4.2. *Pseudo-random numbers*

When random numbers are used in algorithms and crypto-systems, it is essential that the random bits used are unbiased and independent. The speed or the reliability of the algorithm, and the security of the crypto-system depend on the quality of the random numbers used. Natural sources of randomness, such as coins or noise diodes, are fairly slow in generating random bits. On the other hand, for both of these applications, truly random bits could be replaced by any sequence of bits that no polynomial-time algorithm can distinguish from truly random bits. A *pseudo-random number generator* takes a *seed* x, a truly random string of length n, and expands it to a *pseudo-random string* y of length n^k for some constant k. A pseudo-random number generator can be subjected to certain statistical tests. It passes the *next bit test* if after seeing the first i bits of its output y for some $i < n^k$ no polynomial-time algorithm can predict the next bit with probability more than $\frac{1}{2} + n^{-c}$ for any constant c.

Most computers have built-in pseudo-random number generators; one of the simplest ones is the linear congruential generator (where the seed consists of integers a, b, m and x_0, and the pseudo-random numbers are generated by the recurrence $x_{i+1} = ax_i + b \pmod{m}$). This is easily shown to fail the next bit test. Other, more sophisticated pseudo-random number generators can also be shown to output inappropriate sequences, by clever use of the LLL basis reduction algorithm. An example is the binary expansion of algebraic numbers (where the seed is the polynomial defining the algebraic number).

The first provably secure pseudo-random bit generator was developed by Blum and Micali (1984). They proved that pseudo-random bit generators exist, based on the following paradigm: there exist a polynomial-time computable permutation F of the set $\{0,1\}^n$ and a function $B : \{0,1\}^n \to \{0,1\}$, such that $B(x)$ yields a secure bit (as discussed above), but given $F^{-1}(x)$, $B(x)$ can be computed in polynomial time. Such an F is necessarily a one-way function; B is called the "hard core" of F. The Goldreich–Levin construction of a secure bit can be used to show that such a pair of functions exists if a one-way function exists, by taking $F(x) = f(x') \cdot x''$ for some length-preserving one-way function f.

The Blum–Micali pseudo-random number generator produces the sequence $b_0, \ldots, b_k$, defined by $b_i = B(x_{k-i})$, and $x_{i+1} = F(x_i)$, where the random seed is used to select the functions used and the initial vector x_0. The defined pseudo-random number generator can be proved to pass the next bit test. (Informally, suppose that we have an algorithm that can predict $b_{i+1} = B(x_{k-i-1})$, given $b_0, \ldots, b_i$; we will use this to contradict the fact that B yields a secure bit. Note that given just x_{k-i-1}, we can compute $x_{k-i}, \ldots, x_{k-1}$, and use these values, $F^{-1}(x_{k-i}), \ldots, F^{-1}(x_k)$, to compute the bits $b_0, \ldots, b_i$ in polynomial time; hence we can use the assumed procedure to predict b_{i+1}, which is impossible.)

One might wonder whether certain pseudo-random number generators pass statistical tests other than the next bit test. However, Yao (1982) proved that if a pseudo-random number generator passes the next bit test then it passes any statistical test, i.e., no randomized polynomial-time algorithm can distinguish the generated pseudo-random numbers from truly random numbers.

4.3. Zero-knowledge proofs

Let us return to our example with the bank account access. The programmer of the computer of the bank may be tricky and store your password after you have used it once. Can you avoid using it at all and only *prove* to your bank that you have a password (i.e., know a Hamiltonian circuit in the graph G) without giving any help to find it (or mimicking you in any other way)?

This question also comes up in some of the above cryptographic applications: it might be useful to be able to convince someone that a number is the product of two primes, without telling the two primes themselves. This is impossible in the classical sense of proofs, but interactive proof systems, discussed in chapter 29, make it possible. In fact, this was one of the motivating examples for Goldwasser et al. (1989) when developing their notion of interactive proof systems. Informally, an interactive proof of a statement is said to be a zero-knowledge proof if the verifier cannot learn anything from the proof except the validity of the statement.

Before formalizing the notion of zero-knowledge proofs, let us describe a solution to the bank problem (due essentially to M. Blum). To make it more transparent, we imagine another setup: suppose that you are giving a talk on Hamiltonian graphs, and you show your audience a Hamiltonian graph G. For didactical purposes, you want to convince them that the graph G has a Hamiltonian circuit without showing them the circuit itself. This seemingly impossible task can be accomplished using an overhead projector. You prepare two transparencies: both show the same set $V(G)$ of nodes, in some random position; the first shows the edges of the Hamiltonian circuit C in G, the second, the remaining edges of G. On this second transparency, the labels of the nodes are also shown, but not on the first! You place both transparencies on the projector and cover them with a piece of paper, then switch on the projector and let the readers choose whether the top sheet or the top two sheets should be removed.

If only the top sheet is removed, the audience sees the graph G, politely labelled so that the audience can verify that it is indeed the graph G shown at the beginning. If both upper sheets are removed, the audience sees a Hamiltonian circuit on $|V(G)|$ randomly placed nodes in the plane. In either case, no information is given on how the Hamiltonian circuit lies in the graph G. (The only information the audience gets is that they see what they expect.)

On the other hand, if you want to cheat and show a graph G that is not Hamiltonian, then either your bottom transparency does not show a Hamiltonian circuit with the right number of nodes, or the two transparencies together do not show the right graph. So there is a chance of $\frac{1}{2}$ that you get caught! If you repeat this 100 times (a bit boring for a talk, but easily done on paper), and you do not get caught then the audience can be reasonably certain that G is Hamiltonian: your chance of getting away with a non-Hamiltonian graph is one in 2^{100}.

To make the above protocol precise, we have to get rid of the physical devices like projectors and transparencies; but this can be done using the methods of cryptography discussed above. The basic cryptographic tool needed for this is a secure bit-encryption scheme. You must be able to encrypt a bit, so that the audience has no chance to figure out on his own what the bit is, but later you can prove which bit was encrypted. To convince the audience that G has a Hamiltonian circuit, you choose a random permutation P, and use this permutation to obtain a random isomorphic copy G' of the graph G. You encode the permutation, and the

$n(n-1)/2$ bits representing the adjacency matrix of the graph G'. The audience can choose either to ask for a proof that the encoded graph G' is isomorphic to the original, or to ask for a proof that G' has a Hamiltonian circuit. In the first case, you decrypt every encrypted bit, thereby providing the permutation P and the graph G'. In the second case, you decrypt only the bits corresponding to edges participating in the Hamiltonian circuit C.

There are several ways to formalize the notion of zero-knowledge proofs. The one we shall use is *computational* zero knowledge. We say that an interactive protocol is a *zero-knowledge protocol* if the verifier can generate, in randomized polynomial time, a sequence of communication whose distribution is indistinguishable in polynomial time from the distribution of the true transcript of the conversation.

In our example above, the audience could predict that if it chooses to see both transparencies together, it will see the given graph with nodes randomly drawn in the plane; while if it chooses to see the bottom transparency, then it will see a circuit with the right number of (unlabelled) nodes, again randomly drawn in the plane.

In contrast, consider the example of the interactive proof for the graph non-isomorphism problem (chapter 29, section 2). At first sight, this seems to be a zero-knowledge protocol, since, so long as the verifier does not deviate from the protocol, he always knows the prover's next move, and hence could generate the conversation himself. There are zero-knowledge protocols for the graph non-isomorphism problem, but this protocol is not, in fact, zero-knowledge, since the verifier can use it to test if a third graph is isomorphic to one of the two in the input (i.e., the verifier can gain extra information by deviating from the protocol). Goldreich et al. (1986) and, subsequently but independently (under a somewhat stronger assumption), Brassard et al. (1988) proved the following result.

Theorem 4.1. *If one-way functions exist, then every language in $\mathcal{NP}$ has a zero-knowledge interactive proof.*

Proof. To prove that all languages in $\mathcal{NP}$ have zero-knowledge proofs, one merely has to provide a zero-knowledge proof for a single $\mathcal{NP}$-complete problem. We have sketched such a protocol for the Hamiltonian circuit problem. □

Acknowledgments

We would like to thank Shafi Goldwasser and Avi Wigderson for their helpful suggestions and comments on an earlier draft. The research of the first author was supported by grant #1812 of the Hungarian National Research Fund. The research of the second author was supported in part by an NSF Presidential Young Investigator award CCR-89-96272 with matching support from IBM, UPS and Sun and by Air Force Contract AFOSR-86-0078. The research of the third author was supported in part by Air Force Contract AFOSR-86-0078 and by a Packard Research Fellowship. The second and third authors were both supported in part by the National Science Foundation, the Air Force Office of Scientific Research, and the Office of Naval Research, through NSF grant DMS-8920550.

References

Aho, A.V., J.E. Hopcroft and J.D. Ullman

[1983a] *Data Structures and Algorithms* (Addison-Wesley, Reading, MA).

Aho, A.V., J.D. Ullman and M. Yannakakis

[1983b] On notions of information transfer in VLSI circuits, in: *Proc. 15th Annu. ACM Symp. on Theory of Computing* (ACM Press, New York) pp. 133–139.

Ajtai, M.

[1983] Σ_1^1-formulae on finite structures, *Ann. Pure Appl. Logic* **24**, 1–48.

Alon, N., and R.B. Boppana

[1987] The monotone circuit complexity of Boolean functions, *Combinatorica* **7**, 1–22.

Andreev, A.E.

[1985] On a method for obtaining lower bounds for the complexity of individual monotone functions, *Dokl. Akad. Nauk. SSSR* **282**, 1033–1037 [*Soviet Math. Dokl.* **31**, 530–534].

Babai, L., P. Frankl and J. Simon

[1986] Complexity classes in communication complexity theory, in: *Proc. 27th Annu. IEEE Symp. on Foundations of Computer Science* (IEEE Computer Society Press, Washington, DC) pp. 337–347.

Babai, L., N. Nisan and M. Szegedy

[1989] Multiparty protocols and logspace-hard pseudorandom sequences, in: *Proc. 21st Annu. ACM Symp. on Theory of Computing* (ACM Press, New York) pp. 1–11.

Blum, M., and S. Micali

[1984] How to generate cryptographically strong sequences of pseudo-random bits, *SIAM J. Comput.* **13**, 850–864.

Brassard, G., D. Chaum and C. Crépeau

[1988] Minimum disclosure proofs of knowledge, *J. Comput. System Sci.* **37**, 156–189.

Chor, B., and O. Goldreich

[1988] Unbiased bits from sources of weak randomness and probabilistic communication complexity, *SIAM J. Comput.* **17**, 230–261.

Diffie, W., and M.E. Hellman

[1976] New directions in cryptography, *IEEE Trans. Inform. Theory* **IT-22**, 644–654.

Fredman, M.L., and R.E. Tarjan

[1987] Fibonacci heaps and their uses in improved network optimization algorithms, *J. Assoc. Comput. Mach.* **34**, 596 615.

Freivalds, R.

[1979] Fast probabilistic algorithms, in: *Mathematical Foundations of Computer Science 1979, Lecture Notes in Computer Science*, Vol. 74, ed. J. Bečvář (Springer Berlin) pp. 57–69.

Furst, M., J.B. Saxe and M. Sipser

[1981] Parity, circuits, and the polynomial time hierarchy, *Math. Systems Theory* **17**, 13–27.

Goldreich, O.

[1988] Randomness, interactive proofs, and zero-knowledge – a survey, in: *The Universal Turing Machine: A Half-Century Survey*, ed. R. Herken (Kammerer & Unverzagt, Hamburg) 377–405.

Goldreich, O., and L.A. Levin

[1989] A hard-core predicate for all one-way functions, in: *Proc. 21st Annu. ACM Symp. on Theory of Computing* (ACM Press, New York) pp. 25–32.

Goldreich, O., S. Micali and A. Wigderson

[1986] Proofs that yield nothing but their validity and a methodology of cryptographic protocol design, in: *Proc. 27th Annu. IEEE Symp. on Foundations of Computer Science* (IEEE Computer Society Press, Washington, DC) pp. 174–187.

Goldwasser, S.

[1989] Interactive proof systems, in: *Computational Complexity Theory*, ed. J. Hartmanis, *Amer. Math. Soc. Symp. Appl. Math.* **38**, 108–128.

Goldwasser, S., and S. Micali

[1984] Probabilistic encryption, *J. Comput. System Sci.* **28**, 270–299.

Goldwasser, S., S. Micali and C. Rackoff

[1989] The knowledge complexity of interactive proof systems, *SIAM J. Comput.* **18**, 186–208.

Gonnet, G.H.

[1984] *Handbook of Algorithms and Data Structures* (Addison-Wesley, London).

Grötschel, M., L. Lovász and A. Schrijver

[1981] The ellipsoid method and its consequences in combinatorial optimization, *Combinatorica* **1**, 169–197.

Hajnal, A., W. Maass and G. Turán

[1988] On the communication complexity of graph properties, in: *Proc. 20th Annu. ACM Symp. on Theory of Computing* (ACM Press, New York) pp. 186–191.

Halstenberg, B., and R. Reischuk

[1988] On different modes of communication, in: *Proc. 20th Annu. ACM Symp. on Theory of Computing* (ACM Press, New York) pp. 162–172.

Hastad, J.

[1989] Almost optimal lower bounds for small depth circuits, in: *Randomness and Computation*, ed. S. Micali, *Advances in Computing Research*, Vol. 5 (JAI Press, Greenwich, CT) pp. 143–170.

Kalyanasundaram, B., and G. Schnittger

[1987] The probabilistic communication complexity of set intersection, in: *Proc. 2nd Annu. Conf. on Structure in Compl. Theory* (IEEE Computer Society Press, New York) pp. 41–49.

Karchmer, M., and A. Wigderson

[1990] Monotone circuits for connectivity require superlogarithmic depth, *SIAM J. Discrete Math.* **3**, 255–265.

Levin, L.A.

[1986] Average case complete problems, *SIAM J. Comput.* **15**, 285–286.

Lindström, B.

[1969] Determinants on semilattices, *Proc. Amer. Math. Soc.* **20**, 207–208.

Lipton, R.J., and R. Sedgewick

[1981] Lower bounds for VLSI, in: *Proc. 13th Annu. ACM Symp. on Theory of Computing* (ACM Press, New York) pp. 300–307.

Lovász, L.

[1990] Communication complexity: A survey, in: *Paths, Flows, and VLSI, Algorithms and Combinatorics*, Vol. 8, eds. B. Korte, L. Lovász and A. Schrijver (Springer, Berlin) pp. 235–265.

Lovász, L., and M. Saks

[1993] Communication complexity and combinatorial lattice theory, *J. Comput. System Sci.* **47**, 322–349.

Mehlhorn, K., and E.M. Schmidt

[1982] Las Vegas is better than determinism in VLSI and distributed computing, in: *Proc. 14th Annu. ACM Symp. on Theory of Computing* (ACM Press, New York) pp. 330–337.

Merkle, R.C., and M. Hellman

[1978] Hiding information and signatures in trapdoor knapsacks, *IEEE Trans. Inform. Theory* **IT-24**, 525–530.

Pratt, V.R.

[1975] The power of negative thinking in multiplying Boolean matrices, *SIAM J. Comput.* **4**, 326–330.

Rabin, M.O.

[1979] *Digitalized Signatures as Intractable as Factorization*, TR-212 (Laboratory for Computer Science, MIT, Cambridge, MA).

Razborov, A.A.

[1985a] Lower bounds on the monotone circuit complexity of some Boolean functions, *Dokl. Akad. Nauk SSSR* **281**, 798–801 [*Soviet Math. Dokl.* **31**, 354–357].

[1985b] A lower bound on the monotone network complexity of the logical permanent, *Math. Zametki* **37**, 887–900 [*Math. Notes Acad. Sci. USSR* **37**, 485–493].

[1987] Lower bounds on the size of bounded depth circuits over a complete basis with logical addition, *Math. Zametki* **41**, 598–607 [*Math. Notes Acad. Sci. USSR* **41**, 333–338].

Rivest, R.L.

[1990] Cryptography, in: *The Handbook of Theoretical Computer Science*, Vol. A, *Algorithms and Complexity*, ed. J. van Leeuwen (Elsevier, Amsterdam) pp. 717–755.

Rivest, R.L., A. Shamir and L. Adleman

[1978] A method for obtaining digital signatures and public-key cryptosystems, *Comm. ACM* **21**, 120–126.

Shamir, A.

[1982] A polynomial time algorithm for breaking the basic Merkle–Hellman cryptosystem, in: *Proc. 23th Annu. IEEE Symp. on Foundations of Computer Science* (IEEE Computer Society Press, Washington, DC) pp. 145–152.

Smolensky, R.

[1987] Algebraic methods in the theory of lower bounds for Boolean circuit complexity, in: *Proc. 19th Annu. ACM Symp. on Theory of Computing* (ACM Press, New York) pp. 77–82.

Spieker, B., and R. Raz

[1994] On the "log route" conjecture in communication complexity, *Combinatorica*, to appear.

Strassen, V.

[1969] Gaussian elimination is not optimal, *Numer. Math.* **13**, 354–356.

Tardos, É.

[1988] The gap between monotone and non-monotone circuit complexity is exponential, *Combinatorica* **8**, 141–142.

Tarjan, R.E.

[1975] Efficiency of a good but not linear set union algorithm, *J. Assoc. Comput. Mach.* **22**, 215–225.

[1983] *Data Structures and Network Algorithms, CBMS-NSF Regional Conf. Ser. Appl. Math.*, Vol. 44 (SIAM, Philadelphia, PA).

Wilf, H.S.

[1968] Hadamard determinants, Möbius functions and the chromatic number of a graph, *Bull. Amer Math. Soc.* **74**, 960–964.

Yannakakis, M.

[1988] Expressing combinatorial optimization problems by linear programs, in: *Proc. 20th Annu. ACM Symp. on Theory of Computing* (ACM Press, New York) pp. 223–228.

Yao, A.C.

[1982] Theory and applications of trapdoor functions, in: *Proc. 23th Annu. IEEE Symp. on Foundations of Computer Science* (IEEE Computer Society Press, Washington, DC) pp. 80–91.

[1983] Lower bounds by probabilistic arguments, in: *Proc. 24th Annu. IEEE Symp. on Foundations of Computer Science* (IEEE Computer Society Press, Washington, DC) pp. 420–428.

Yao, A.C.-C.

[1985] Separating the polynomial-time hierarchy by oracles, in: *Proc. 26th Annu. IEEE Symp. on Foundations of Computer Science* (IEEE Computer Society Press, Washington, DC) pp. 1–10.

CHAPTER 41

Combinatorics in Pure Mathematics

L. LOVÁSZ

Department of Computer Science, Yale University, New Haven, CT 06520, USA

L. PYBER

Mathematical Research Institute, Hungarian Academy of Sciences, Budapest, Hungary H-1053

D.J.A. WELSH

Merton College and Mathematical Institute, University of Oxford, Oxford OX1 4JD, UK

G.M. ZIEGLER

Konrad-Zuse-Zentrum für Informationstechnik (ZIB), Heilbronner Str. 10, D-10711 Berlin, Germany

Contents

HANDBOOK OF COMBINATORICS
Edited by R. Graham, M. Grötschel and L. Lovász

It is a beautiful feature of mathematics that quite often results and methods from one branch can be applied to solve problems in a seemingly distant branch in a successful and often surprising way. In fact, techniques from classical fields like analysis and linear algebra belong to the toolbox of every mathematician. In virtually no field would one be surprised to see a generating function or the computation of the eigenvalues of a matrix. This Handbook contains several chapters showing how algebra, topology, probability theory etc. can be applied to combinatorial problems in a deep way.

The application of combinatorial methods in other areas is not so common (this may be due to the relative youth of the subject). This chapter contains a collection of examples showing the application of a variety of combinatorial ideas to other areas.

In some cases, it is a specific combinatorial theorem that is used. For example, in section 1, theorems from extremal combinatorics are applied in the theory of finite dimensional Banach spaces, to prove embeddability and near-embeddability results. In section 3, the Marriage theorem is applied to give a very simple constructive proof for the existence of the Haar measure on compact topological groups.

Elsewhere, it is a combinatorial structure whose presence should be recognized to shed light on a seemingly hopelessly complicated situation. Section 4 illustrates the use of graphs and coherent configurations in the study of permutation groups. Finding the combinatorial structure which retains just enough of the group structure is the key point. Section 5 describes a method of analysing a commutative ring (like, for example, the coordinate ring of a Grassmann or Schubert variety) by writing it as an *algebra with straightening law* over a finite poset, and then utilizing poset combinatorics to identify algebraic structure of the ring and to prove fundamental algebraic properties of some classical projective varieties. Section 6 shows how the algebraic and topological structure of a *complex hyperplane arrangement* is determined by the intersection lattice of the arrangement (i.e., by a matroid).

And in some cases, problems with a very classical flavour turn out to be so closely related to combinatorial problems that even the direction of the interaction is difficult to tell. In section 2 we see how embeddings of finite metric spaces in classical Banach spaces are related to "hypermetric inequalities", to multicommodity flows, and to lattice-point-free ellipsoids. In section 7 we sketch how a recent topological invariant of knots, the Jones polynomial, can be derived from the Tutte polynomial of an appropriate graph, and how this connection can be used to settle a long-standing conjecture in the theory of knots.

1. Sections of finite dimensional Banach spaces

It is well known that if N_1 and N_2 are two norms on a finite dimensional real linear space $\mathbb{R}^n$ then there exist constants $c_1, c_2 > 0$ such that

$$c_1 N_1(x) \leqslant N_2(x) \leqslant c_2 N_1(x)$$

for all vectors x. These constants depend on the norms; but if we allow a linear transformation, then any given norm N_1 can be transformed into one which is more similar to another given norm N_2. In fact, let $K = K(N_1) = \{x \in \mathbb{R}^n: N_1(x) \leqslant 1\}$ be the unit ball of the first norm; this is a convex body centrally symmetric with respect to the origin. It can be shown that there exists an ellipsoid E inscribed in K with maximum volume and it is unique; E is called the inscribed Löwner–John ellipsoid of K (cf. chapters 19 and 30, and also Grötschel et al. 1988). Perhaps the most important property of E is that if we blow up E by a factor of $\sqrt{n}$ then the resulting ellipsoid will contain K (for this result, the central symmetry of K is needed; for general convex bodies, $\sqrt{n}$ has to be replaced by n).

If we apply a linear transformation that maps E onto the unit ball, then N_1 will be transformed into a norm N_1' satisfying

$$\frac{1}{\sqrt{n}}\|x\| \leqslant N_1'(x) \leqslant \|x\|$$

(here $\|x\|$ denotes the Euclidean norm). Applying this argument to the other norm as well, we get that N_1 can be transformed by a linear transformation into a norm N_1'' such that

$$N_1''(x) \leqslant N_2(x) \leqslant nN_1''(x)$$

for all $x \in \mathbb{R}^n$.

One can achieve a better approximation if one is allowed to restrict the norms to appropriate subspaces; this way one obtains theorems with the flavor of Ramsey's theorem (see chapter 25). A classical result of Dvoretzky (1959) asserts that *for every positive integer k and every $\varepsilon > 0$ there exists an $n = n(k, \varepsilon) > 0$ such that for every norm N on $\mathbb{R}^n$ there exists a constant $C > 0$ and a subspace V of $\mathbb{R}^n$ with dimension k such that for every $x \in V$,*

$$C \cdot \|x\| \leqslant N(x) \leqslant (1 + \varepsilon)C \cdot \|x\|.$$

The following result of Figiel et al. (1977) illustrates how to obtain finer measures of the approximation of an arbitrary norm by the Euclidean norm on a subspace. Let $M_2 = M_2(N)$ denote the square root of the average of $N(x)^2$ over all vectors with unit Euclidean length. Then *for every $\varepsilon > 0$ there exists a $\delta > 0$ such that for every norm N on $\mathbb{R}^n$ such that $N(x) \leqslant \|x\|$ there exists a subspace V with dimension at least $\delta n M_2^2$ such that $N(x) \geqslant (1 - \varepsilon)\|x\|$ for every $x \in V$.*

The proof of such results on the ℓ_2 norm is geometric. Our main topic will be two results with a similar flavour concerning the ℓ_1-norm $\|x\|_1 = \sum_i \|x_i\|$ and the ℓ_∞-norm $\|x\|_\infty = \max_i \|x_i\|$. Both proofs use a number of combinatorial tools.

We introduce some quantities analogous to the quantity $M_2(N)$ defined above. Let N be an arbitrary norm on $\mathbb{R}^n$. Let $M_1 = M_1(N)$ denote the average of $N(x)$ over all ±1-vectors, and $M_\infty = M_\infty(N)$, the maximum of $N(x)$ over all ±1 vectors. If we have $N(e_i) = 1$ for all i then $1 \leqslant M_1(N) \leqslant n$, with equality in the first inequality iff $N = \|\cdot\|_\infty$ and equality in the second inequality iff $N = \|\cdot\|_1$.

Note that $N(x)$ is a convex function and hence we could take the maximum over the whole cube spanned by $\{-1,1\}^n$ (the unit ball of $\|\cdot\|_\infty$) instead. So we have

$$N(x) \leqslant M_\infty \|x\|_\infty$$

for all x.

We can also formulate an opposite inequality. Assume that $N(e_i) = 1$ for $i = 1,\ldots,n$. Let $b_i^{\mathrm{T}}x = 1$ be the equation of a supporting hyperplane to K at the point e_i (so b_i is in the polar body). Set $b_i = (b_{i1},\ldots,b_{in})^{\mathrm{T}}$; then $b_{ii} = 1$. Note that $M_\infty \geqslant \max_i \sum_j |b_{ij}|$. Let

$$M_0 = \min_i \min_{\substack{x \in \{-1,1\}^n \\ x_i = 1}} b_i^{\mathrm{T}}x = \min_i \Big\{ 1 - \sum_{i \neq j} |b_{ij}| \Big\}$$

(this may be negative, which will be a trivial case). For any vector $x \in \mathbb{R}^n$, we have

$$N(x) = \max\{b^{\mathrm{T}}x:\ b \in K^*\}.$$

Let i be the index with $|x_i|$ maximum, and assume that, say, $x_i \geqslant 0$. Then

$$N(x) \geqslant b_i^{\mathrm{T}}x = \sum_j x_j b_{ij} \geqslant x_i - \sum_{j \neq i} |x_j| \cdot |b_{ij}| \geqslant x_i \Big(1 - \sum_{j \neq i} |b_{ij}|\Big) = M_0 \|x\|_\infty.$$

Our first topic is a theorem of Milman (1982). Let N be a norm in $\mathbb{R}^n$ whose unit ball is a polyhedron (this is so far not a very severe restriction since any norm can be approximated arbitrarily well by a polyhedral norm). Then N can be written as $N(x) = \|Ax\|_\infty$ with an appropriate matrix $A: \mathbb{R}^m \to \mathbb{R}^n$. Now assume that every entry of A is 1 or -1. Then A has at most 2^n distinct rows. Observe that if the number of rows is exactly 2^n then $N(x) = \|x\|_1$.

Theorem 1.1. *If $N(x) = \|Ax\|_\infty$, where A is a ± 1 matrix with m (distinct) rows and n columns, then there exists a linear subspace $V \subseteq \mathbb{R}^n$ with $\dim V \geqslant \lfloor \ln m / \ln n \rfloor$ such that the restriction of N to V is isometric to an ℓ_1 norm.*

Proof. Let $k = \lfloor \ln m / \ln n \rfloor$; we may assume that $k > 1$. Then $m \geqslant n^k > 1 + n + \binom{n}{2} + \cdots + \binom{n}{k}$, and hence by the Sauer–Shelah theorem (chapter 24, Theorem 4.8), we can choose k columns of A (say, the first k columns) such that every ± 1-vector arises as the restriction of some row of A to these columns. Let V be the subspace of $\mathbb{R}^n$ consisting of those vectors whose last $n-k$ coordinates are 0. Then for every $v \in V$,

$$N(v) = \|Av\|_\infty = \max_i \Big| \sum_{j=1}^{k} a_{ij} v_j \Big| = \sum_{j=1}^{k} |v_j| = \|v\|_1,$$

since the maximum is attained for the row of A for which $a_{ij} = \operatorname{sign} v_j$. □

A corollary of this theorem uses the quantity M_1 to bound the dimension of V:

Corollary 1.2. *If $N(x) = \|Ax\|_\infty$, where A is a ± 1-matrix, then there exists a linear subspace $V \subseteq \mathbb{R}^n$ with $\dim V \geqslant M_1^2/(2n \ln n)$ such that the restriction of N to V is isometric to an ℓ_1 norm.*

Proof. Assume that A has m distinct rows; trivially, $m \geqslant n$. We show that $M_1 < \sqrt{2n \ln m}$; this then implies that the subspace V in the assertion of Theorem 1.1 satisfies $\dim V > M_1^2/(2n \ln n)$.

We want to bound, for a random ± 1 vector w, the expectation of $\|Aw\|_\infty$. Let $a_1, \ldots, a_m$ be the rows of A. Then $a_i w$ is the sum of n independent random variables assuming 1 and -1 with equal probability, and hence by Chernoff's inequality (see chapter 33),

$$\mathrm{Prob}\left(|a_i w| > \sqrt{n \ln m}\right) < \mathrm{e}^{-2 \ln m} = \frac{1}{m^2}.$$

Hence

$$\mathrm{Prob}\left(\max_i |a_i w| > \sqrt{n \ln m}\right) < m \cdot \mathrm{e}^{-2 \ln m} = \frac{1}{m},$$

and so the expectation of $\|Aw\|_\infty$ is at most $(1 - 1/m)\sqrt{n \ln m} + n/m < \sqrt{2n \ln m}$. (This argument is essentially the same as the probabilistic upper bound on the discrepancy of a hypergraph with n vertices and m edges; cf. chapter 26.) □

The next result whose proof uses combinatorial tools is due to Alon and Milman (1983).

Theorem 1.3. *Let N be any norm on $\mathbb{R}^n$ such that $N(e_i) = 1$ for all i.*

(i) *There exists a subspace V spanned by $\lfloor \sqrt{n} \rfloor$ basis vectors e_i such that $M_\infty(N|_V) \leqslant 8M_1(N)$; in other words we have, for all $x \in V$,*

$$N(x) \leqslant 8M_1 \|x\|_\infty.$$

(ii) *For every $\varepsilon > 0$ there exists a subspace V spanned by $\lfloor \varepsilon n/(8M_\infty) \rfloor$ basis vectors e_i such that $M_0(N|_V) \geqslant 1 - \varepsilon$, and hence we have for all $x \in V$,*

$$(1 - \varepsilon)\|x\|_\infty \leqslant N(x).$$

Combining (i) with (ii) we obtain that there exists a subspace V spanned by $\lfloor \sqrt{n}/(128M_1) \rfloor$ basis vectors e_i such that for all $x \in V$,

$$\tfrac{1}{2}\|x\|_\infty \leqslant N(x) \leqslant 8M_1 \|x\|_\infty.$$

Proof. (i) This part of the proof uses a combinatorial lemma which is analogous to the Sauer–Shelah Theorem. Let $h(n,k)$ denote the smallest integer for which the following holds: whenever $R \subseteq \{-1,1\}^n$ with $|R| > h(n,k)$, then there exists a subset $I \subseteq \{1, \ldots, n\}$ with $|I| = k$ such that for every $q \in \{-1,1\}^S$ there exist two vectors $u, w \in R$ such that $u_i = w_i = q_i$ if $i \in I$ and $u_i = -w_i$ if $i \notin I$.

Lemma 1.4.

$$h(n,k)=\begin{cases}\displaystyle\sum_{i=0}^{(n+k-1)/2}\binom{n}{i}, & \textit{if } n+k \textit{ is odd,}\\[2ex] \displaystyle\binom{n-1}{(n+k)/2}+\sum_{i=0}^{(n+k-2)/2}\binom{n}{i}, & \textit{if } n+k \textit{ is even.}\end{cases}$$

For a proof of this lemma, which uses the "down-shift" technique (cf. chapter 24), we refer to the original paper.

Now let R denote the set of vectors $w\in\{-1,1\}^n$ such that $N(w)\leqslant 8M_1$. Let k be the least integer such that $k\geqslant\sqrt{n}$ and $k\equiv n-1 \pmod 2$. Then

$$M_1=2^{-n}\sum_{w\in\{-1,1\}^n}N(w)\leqslant 2^{-n}(2^n-|R|)(8M_1),$$

whence (if n is large enough)

$$|R|\geqslant\tfrac{7}{8}2^n>\sum_{i=0}^{(n+k-1)/2}\binom{n}{i}.$$

Hence by Lemma 1.4, there exists a subset $I\subseteq\{1,\dots,n\}$ such that $|I|=k$ and for every $q\in\{-1,1\}^S$ there exist two vectors $u,w\in R$ such that $u_i=w_i=q_i$ if $i\in I$ and $u_i=-w_i$ if $i\notin I$. We claim that the subspace V spanned by $\{e_i\colon i\in I\}$ satisfies the requirement in (i).

By our discussion of M_∞, it suffices to show that for every ±1-vector v spanned by e_i $(i\in I)$ we have $N(q)\leqslant 8M_1$. To this end, let q be the restriction of v to the coordinates in I and consider the vectors $u,w\in R$ as above. Then $v=(1/2)(u+w)$ and hence

$$N(v)\leqslant\tfrac{1}{2}\big(N(u)+N(w)\big)\leqslant 8M_1.$$

(ii) The proof of this second part also uses a lemma with a combinatorial flavor (Johnson and Schechtman 1981):

Lemma 1.5. *Let A be an $n\times n$ matrix with non-negative entries and with 0's in the diagonal. Assume that each row-sum of A is at most D. Then for every $k\geqslant 1$, A has a $k\times k$ principal submatrix A' such that each row-sum of A' is at most $8kD/n$.*

This lemma can be cast in graph-theoretic terms: *if G is a directed graph with maximum outdegree D and $k\geqslant 1$, then G contains an induced subgraph on k nodes with maximum outdegree at most* $8kD/n$. For undirected graphs, this follows from a theorem of Lovász asserting that *if G is an undirected graph with maximum degree D and $t\geqslant 1$ then the nodes of G can be partitioned into $\lceil(D-1)/t\rceil$ classes such that each class induces a subgraph with maximum degree at most t* (cf. chapter 4, Theorem 4.2). The directed version follows by deleting the nodes with indegree larger than $2D$ (this means at most half of the nodes), and then applying the undirected version to the remaining graph, disregarding the orientation.

Return to the proof of Theorem 1.3. Let $k = \lfloor \varepsilon n/(8M_\infty) \rfloor$. With the notation introduced at the beginning of the section, consider the matrix $P = (p_{ij})$ defined by

$$p_{ij} = \begin{cases} 0, & \text{if } i = j, \\ |b_{ij}|, & \text{if } i \neq j. \end{cases}$$

Then every row-sum of P is bounded by M_∞:

$$\sum_j p_{ij} \leqslant \sum_j |b_{ij}| \leqslant M_\infty.$$

Hence by Lemma 1.4, A has a $k \times k$ principal submatrix in which every row-sum is at most ε. Let, say, the upper left $k \times k$ submatrix of P have this property, and let V be the subspace spanned by $e_1, \ldots, e_k$. Then

$$M_0(N|_V) = \min_{i \leqslant k} \Bigl(1 - \sum_{\substack{j \leqslant k \\ j \neq i}} |b_{ij}| \Bigr) \geqslant 1 - \varepsilon,$$

and hence for every $x \in V$,

$$N(x) \geqslant (1 - \varepsilon) \|x\|_\infty. \qquad \square$$

We remark that by a rather standard "partitioning" argument one can improve the upper bound in (i) at the cost of decreasing the dimension of the subspace, and prove the following:

Corollary 1.6. *For every norm N on $\mathbb{R}^n$ and every $\varepsilon > 0$ there exists a linear embedding $A : \mathbb{R}^k \to \mathbb{R}^n$ where $k = n^{\varepsilon/(4 \ln M_1)}$ such that for every $x \in \mathbb{R}^k$,*

$$\|x\|_\infty \leqslant N(Ax) \leqslant (1 + \varepsilon) \|x\|_\infty.$$

2. Embeddings of finite metric spaces and hypermetric inequalities

Let us start by recalling the following classical result of Cayley:

Theorem 2.1. *A symmetric matrix $D = (d_{ij})_{i=1}^{n}{}_{j=1}^{n}$ is the matrix of mutual distances of n (not necessarily distinct) points in $\mathbb{R}^n$ (or in the Hilbert space) if and only if $D \geqslant 0$, $d_{ii} = 0$ for all i, and the matrix $(a_{ij})_{i,j=1}^{n-1}$ is positive semidefinite, where*

$$a_{ij} = d_{in}^2 + d_{jn}^2 - d_{ij}^2.$$

The proof of this theorem is rather straightforward linear algebra. Another way to state this condition is that the matrix $D^{(2)}$ obtained by squaring each entry of

D is *of negative type*, which means that for every vector $x \in \mathbb{R}^n$ with $\sum_i x_i = 0$, we have

$$\sum_i \sum_j d_{ij}^2 x_i x_j \leqslant 0.$$

We may ask analogous questions about embeddability in other important Banach spaces, such as the L_1 space. The answer to such questions often leads to combinatorial considerations which tie these issues to polyhedral combinatorics, lattice geometry, and flow theory.

So let D be an $n \times n$ matrix and assume that D is the matrix of mutual distances of n points in L_1. There are some obvious conditions that D has to satisfy:

$$d_{ij} = d_{ji}, \tag{2.1}$$

$$d_{ij} \geqslant 0, \tag{2.2}$$

$$d_{ii} = 0, \tag{2.3}$$

and of course the triangle inequality:

$$d_{ij} + d_{jk} \geqslant d_{ik}. \tag{2.4}$$

A matrix satisfying these conditions is called a *metric*. (To be precise, it should be called a semimetric, since in the definition of metric it is usually assumed that distinct points have positive distance; but we allow that two rows of the matrix be represented by the same point, and so it is more convenient to allow $d_{ij} = 0$. Also note that (2.1), (2.3) and (2.4) imply (2.2).) All metrics for a fixed n form a convex cone M_n, which we call the *metric cone*. Since M_n is defined by a finite number of linear inequalities, it is a polyhedral cone.

Now consider the fact that D is a metric that is L_1-embeddable, i.e., it can be represented by a measurable space $(\Omega, \mathscr{A}, \mu)$ with finite measure and by integrable functions $f_1, \ldots, f_n : \Omega \to \mathbb{R}$ so that

$$d_{ij} = \int_\Omega |f_i - f_j| \, \mathrm{d}\mu.$$

It is not difficult to see that these functions can be chosen 0–1-valued and $(\Omega, \mathscr{A}, \mu)$ can be chosen so that it consists of a finite number of atoms. So if D is L_1-embeddable then there always exists a representation in terms of a finite set Ω, a weighting $\mu : \Omega \to \mathbb{R}_+$, and subsets $A_i \subseteq \Omega$ such that

$$d_{ij} = \mu(A_i \triangle A_j).$$

It is easy to verify from this interpretation that for each fixed n, L_1-embeddable metrics form a convex cone: if D is represented by $(\Omega, \mu, A_1, \ldots, A_n)$ and D' is represented by $(\Omega', \mu', A'_1, \ldots, A'_n)$ (where we may assume that $\Omega \cap \Omega' = \emptyset$), then λD is represented by $(\Omega, \lambda\mu, A_1, \ldots, A_n)$ and $D + D'$ is represented by $(\Omega \cup \Omega', \mu \cup \mu', A_1 \cup A'_1, \ldots, A_n \cup A'_n)$. This cone is called the *Hamming cone* and is denoted by H_n. (The fact that L_1-embeddable metrics form a convex cone is not entirely obvious: this is not true, e.g., for L_2-embeddable metrics.)

It should also be noted that every L_2-embeddable matrix is also L_1-embeddable. In fact, assume that D can be represented by the Euclidean distances in a set $\{v_1, \ldots, v_n\} \subset \mathbb{R}^N$. Let Ω be the set of halfspaces in $\mathbb{R}^N$ separating at least one pair $\{v_i, v_j\}$. There exists a translation- and rotation-invariant measure μ on the halfspaces in $\mathbb{R}^N$, and it is easy to see that $\mu(\Omega)$ is finite. Let A_i denote the set of halfspaces in Ω containing v_i; then $\mu(A_i \triangle A_j)$ is proportional to the Euclidean distance of v_i and v_j.

Returning to L_1-embeddability, we want to describe the Hamming cone H_n. The construction showing that it is a cone can be used "backwards" to show that *every Hamming metric is the sum of Hamming metrics on single atoms.* A Hamming metric on a single atom $\Omega = \{a\}$ is quite simple; normalizing by $\mu(a) = 1$, every row must be represented by either $\emptyset$ or $\{a\}$, and so for an appropriate $S \subseteq \{1, \ldots, n\}$, the matrix D looks like

$$d_{ij} = \begin{cases} 1, & \text{if } i \in S \text{ and } j \notin S \text{ or vice versa,} \\ 0, & \text{otherwise.} \end{cases}$$

Such a matrix will be called a cut matrix (or cut metric). Our argument shows that every Hamming metric is a non-negative linear combination of cut metrics. Hence we have:

Proposition 2.2. *The Hamming cone is polyhedral and its extreme rays are spanned by cut metrics.*

In the spirit of polyhedral combinatorics, a next task would be to describe the facets of H_n. Unfortunately, no complete description is available; membership in H_n is NP-hard to decide (Karzanov 1985). The Hamming cone can be viewed as the *cut cone* of the complete graph, and general results on the cut polyhedron yield many facets. Nevertheless, much of the development of the two topics has been independent; see Barahona and Mahjoub (1986), Deza and Laurent (1992a,b,c).

An important class of inequalities with geometric flavour is that D is of negative type, i.e., for every $x \in \mathbb{R}^n$ we have

$$\sum_i x_i = 0 \quad \Longrightarrow \quad \sum_i \sum_j d_{ij} x_i x_j \leqslant 0. \tag{2.5}$$

(for L_2-embeddable metrics, we had the squared distances in a similar inequality). To see that (2.5) holds for every Hamming metric, it suffices to consider the extreme rays of H_n, i.e., the cut metrics; and for such a metric, we have

$$\sum_i \sum_j d_{ij} x_i x_j = 2 \sum_{i \in S} \sum_{j \notin S} x_i x_j = 2 \Big(\sum_{i \in S} x_i \Big) \Big(\sum_{j \notin S} x_j \Big) = -2 \Big(\sum_{i \in S} x_i \Big)^2 \leqslant 0.$$

Another way to state this inequality is the following. Let $i_1, \ldots, i_k, j_1, \ldots, j_k$ be (not necessarily distinct) indices from $\{1, \ldots, n\}$. Then

$$\sum_{1 \leqslant p < r \leqslant k} d(i_p, i_r) + \sum_{1 \leqslant p < r \leqslant k} d(j_p, j_r) \leqslant \sum_{\substack{1 \leqslant p \leqslant k \\ 1 \leqslant r \leqslant k}} d(i_p, j_r) \tag{2.6}$$

(distances between the same kind of points sum up to not more than distances between different kinds of points). To derive (2.6) from (2.5), let x_i be the number of times i occurs among the indices j_r, less the number of times i occurs among the indices i_p; then $\sum_i x_i = 0$ and (2.5) implies (2.6). Conversely, (2.6) implies (2.5) directly for integral vectors x, from which the rational case follows by homogeneity and the general case follows by continuity.

There is a way to sharpen this inequality, which leads to perhaps the most important class of inequalities valid for H_n (Deza 1960, Kelly 1970). Let $i_1, \ldots, i_k, j_1, \ldots, j_{k+1}$ be (not necessarily distinct) indices from $\{1, \ldots, n\}$. Then

$$\sum_{1 \leqslant p < r \leqslant k} d(i_p, i_r) + \sum_{1 \leqslant p < r \leqslant k+1} d(j_p, j_r) \leqslant \sum_{\substack{1 \leqslant p \leqslant k \\ 1 \leqslant r \leqslant k+1}} d(i_p, j_r) \tag{2.7}$$

is valid for every Hamming metric. The triangle inequality is just the special case $k = 1$. This is why these inequalities are called *hypermetric inequalities*.

We can give a linear algebraic formulation of these inequalities analogous to (2.5):

$$\sum_i x_i = 1, \ x_i \text{ integer} \quad \Longrightarrow \quad \sum_i \sum_j d_{ij} x_i x_j \leqslant 0 \tag{2.8}$$

(the condition that the x_i are integral cannot be dropped; else, the inequality would hold for every vector x, and so D would be negative semidefinite, which is impossible for $D \neq 0$ as $\operatorname{trace}(D) = 0$). From this formulation, the inequality can be proved for cut metrics (and hence for all Hamming metrics) similarly to the proof of (2.6) above.

The hypermetric inequalities hold, in particular, for the Euclidean metric (this is not quite obvious to see directly), and have interesting applications in discrete geometry (see, e.g., Kelly 1975).

The convex cone defined by the inequalities (2.5) (or (2.6)) is called the *negative type cone*, while that defined by the inequalities (2.7) (or (2.8)) is called the *hypermetric cone*. The negative type cone is non-polyhedral for $n \geqslant 3$; but Deza et al. (1993) prove the following:

Theorem 2.3. *The hypermetric cone is polyhedral for every $n \geqslant 2$.*

This fact is non-trivial, since seemingly there are infinitely many defining inequalities. To analyze the structure of hypermetric inequalities, let us substitute $x_n = 1 - \sum_{i=1}^{n-1} x_i$ in (2.8): then we get that for every $x \in \mathbb{Z}^{n-1}$,

$$\sum_{i,j \leqslant n-1} (d_{in} + d_{jn} - d_{ij}) x_i x_j - 2 \sum_{i \leqslant n-1} d_{in} x_i \geqslant 0. \tag{2.9}$$

It follows easily that the matrix $A = (d_{in} + d_{jn} - d_{ij})_{i=1}^{n-1}{}_{j=1}^{n-1}$ is positive semidefinite (in fact, this is equivalent to (2.5)). Let E denote the solution set of

$$\sum_{i,j\leqslant n-1} (d_{in} + d_{jn} - d_{ij})x_i x_j - 2 \sum_{i\leqslant n-1} d_{in}x_i \leqslant 0.$$

Property (2.9) implies that E contains no lattice points in its interior. On the other hand, it follows by substitution that the zero vector and the unit vectors are on the boundary of E. If A is positive definite, then E is an ellipsoid; in general, it is a direct product of a linear subspace with an ellipsoid, which we call a *generalized ellipsoid*. So every hypermetric corresponds to a lattice-point-free generalized ellipsoid in $\mathbb{R}^{n-1}$, having 0 and the unit vectors on its boundary. Conversely, every such generalized ellipsoid yields a hypermetric (uniquely up to a scalar factor). Such a generalized ellipsoid corresponds to an extreme ray of the hypermetric cone if and only if the system of hypermetric equalities satisfied by d admit d as a unique solution (up to a scaling).

Deza et al. use this description of the extreme rays, together with a theorem of Voronoi on the finiteness of affine types of Delaunay polytopes of a lattice of a given dimension, to show that the number of extreme rays of the hypermetric cones is finite, i.e., this cone is polyhedral.

An interesting connection between the metric cone and multicommodity flows was pointed out by Avis and Deza (1991). We formulate the multicommodity flow problem as follows (cf. chapter 2). Consider the complete graph K_n on nodes $\{1,\ldots,n\}$. For each unordered pair i,j of nodes, we are given a capacity $c_{ij} \geqslant 0$, and a demand $d_{ij} \geqslant 0$. So (c_{ij}) and (d_{ij}) are symmetric matrices, and we assume that $c_{ii} = d_{ii} = 0$ for each node i. We want to find a flow f_{ij} from i to j of value d_{ij} for every $1 \leqslant i < j \leqslant n$ such that

$$\sum_{i,j} |f_{ij}(uv)| \leqslant c_{uv}$$

for every pair u,v. We say that the pair (C,D) is *feasible* if such a system of flows exists. (If we want to work on a graph different from the complete graph, we can set $c_{uv} = 0$ for the non-adjacent pairs. Similarly, we set $d_{ij} = 0$ if we do not want a flow from i to j.) Now this is a system of linear inequalities and a characterization of feasibility can be obtained from the Farkas lemma. However, this condition is not very transparent; Iri (1970) and Onaga and Kakusho (1971) managed to replace it by the following elegant criterion (chapter 2, Theorem 8.1): *a capacity–demand pair (C,D) is feasible if and only if $C - D$ is contained in the polar of the metric cone.*

A special necessary condition for feasibility, also formulated in chapter 2, is the *cut condition:* for every partition $\{S, V \setminus S\}$, we must have

$$\sum_{i\in S, j\in V\setminus S} d_{ij} \leqslant \sum_{i\in S, j\in V\setminus S} c_{ij}.$$

It is easy to see that this is equivalent to saying that $C - D$ gives a non-negative inner product with every cut metric. Since cut metrics are just the extreme rays of the Hamming cone H_n, this means that $C - D$ is contained in the polar of the Hamming cone. (Note that $H_n \subseteq M_n$ and hence $H_n^* \supseteq M_n^*$.)

Various results on multicommodity flows discussed in chapter 2 are worth rephrasing in terms of these cones (see in particular Theorem 8.6). The Max-Flow–Min-Cut theorem says that if a symmetric matrix A has 0's in its diagonal and has only one pair of negative entries, then $A \in H_n^*$ implies $A \in M_n^*$. Hu's theorem on 2-commodity flows implies that this holds also when A has two pairs of negative entries. Papernov's work can be viewed as a characterization of all patterns of negative entries of a matrix in $H_n^* \setminus M_n^*$. Other results on multicommodity flows exclude certain supports for a matrix in $H_n^* \setminus M_n^*$; e.g., Seymour's theorem 8.6(g) in chapter 2 implies that such a support cannot be the adjacency matrix of a planar graph.

3. Matchings and the Haar measure

A fundamental notion in measure theory is the Haar measure on locally compact topological groups. We shall show that matching theory can be applied to give a simple and constructive proof of the existence of the Haar measure in the compact case (Rota and Harper 1971). We in fact prove the existence of invariant integration; this result, proved by von Neumann, is equivalent to the existence of the Haar measure. We refer to Halmos (1950) for measure-theoretic background. For further applications of matching theory to measure theory, see also Lovász and Plummer (1986).

Let G be a compact topological group (i.e., a group G endowed with a compact topology such that the group operations of multiplication and inverse are continuous). Let $C(G)$ denote the space of real valued continuous functions defined on G. An *invariant integration* is a functional defined on $C(G)$, having the following properties:

(1) $L(\alpha f + \beta g) = \alpha L(f) + \beta L(g)$ (linearity);

(2) if $f \geqslant 0$ then $L(f) \geqslant 0$ (positivity);

(3) if $\mathbf{1}$ denotes the identically 1 function then $L(\mathbf{1}) = 1$ (normalization);

(4) if $s, t \in G$ and $f \in C(G)$, and g is defined by $g(x) = f(sxt)$ then $L(g) = L(f)$ (double translation invariance).

Theorem 3.1. *Every compact topological group admits an invariant integration.*

For the proof, we need some notation. If A is a finite set and $f : A \to \mathbb{R}$ then we set

$$\bar{f}(A) = \frac{1}{|A|} \sum_{a \in A} f(a).$$

If $\mathscr{H} = (V, E)$ is a (finite or infinite) hypergraph and $f : V \to \mathbb{R}$ then we set

$$\delta(f, \mathscr{H}) = \sup\{|f(x) - f(y)| : x, y \in U \text{ for some } U \in \mathscr{H}\}.$$

The combinatorial lemma we use is the following.

Lemma 3.2. *Let $\mathcal{H}$ be a hypergraph and let A, B be two minimum-cardinality blocking sets of $\mathcal{H}$. Assume that A (and B) are finite. Then*

$$|\bar{f}(A)-\bar{f}(B)| \leqslant \delta(f,\mathcal{H}).$$

Proof. Consider the bipartite graph whose color classes are A and B, where $a \in A$ is connected to $b \in B$ iff there exists an edge $U \in \mathcal{H}$ containing both a and b. We claim that this bipartite graph has a perfect matching. We verify the conditions in the Marriage theorem (see chapter 3, Corollary 2.5). It is clear that $|A| = |B|$. Let $T \subseteq A$ and let $N(T)$ denote the set of neighbors of T in B.

We show that the set $A' = A \cup N(T) \setminus T$ is also a blocking set for $\mathcal{H}$. In fact, every edge U of $\mathcal{H}$ meets A as well as B, since A and B are blocking. Let $a \in U \cap A$ and $b \in U \cap B$. If $a \notin T$ then we are done. If $a \in T$ then $b \in N(T)$ and so again A' meets U.

So A' is a blocking set and hence $|A'| \geqslant |A|$, which implies that $|N(T)| \geqslant |T|$. Thus the Marriage theorem applies and G has a perfect matching, say $\{a_1b_1, \ldots, a_nb_n\}$. Then

$$\begin{aligned}|\bar{f}(A)-\bar{f}(B)| &= \left|\frac{1}{n}\sum_{i=1}^{n}(f(a_i)-f(b_i))\right| \\ &\leqslant \frac{1}{n}\sum_{i=1}^{n}|f(a_i)-f(b_i)| \\ &\leqslant \frac{1}{n}n\delta(f,\mathcal{H}) = \delta(f,\mathcal{H}). \qquad \square\end{aligned}$$

Proof of Theorem 3.1. Let U be a non-empty open subset of G. We denote by $\mathcal{H}_U$ the hypergraph with underlying set G, whose edges are the *translates* of U, i.e., the sets $sUt = \{sut\colon u \in U\}$ $(s, t \in G)$. A blocking set of $\mathcal{H}_U$ is called a *U-net*. It follows from the compactness of the group G that there exists a finite U-net. Let $f \in C(G)$; we want to define the value $L(f)$.

Claim. *Let U and V be non-empty open subsets of G. Let A be a minimum-cardinality U-net and B, a minimum-cardinality V-net. Then*

$$|\bar{f}(A)-\bar{f}(B)| \leqslant \delta(f,\mathcal{H}_U)+\delta(f,\mathcal{H}_V).$$

To prove this claim, observe that Ab is also a minimum U-net for any $b \in B$, and hence by Lemma 3.2,

$$|\bar{f}(A)-\bar{f}(Ab)| \leqslant \delta(f,\mathcal{H}_U).$$

Hence

$$\begin{aligned}|\bar{f}(A)-\bar{f}(AB)| &= \left|\bar{f}(A)-\frac{1}{|B|}\sum_{b\in B}\bar{f}(Ab)\right| \\ &\leqslant \frac{1}{|B|}\sum_{b\in B}|\bar{f}(A)-\bar{f}(Ab)| \leqslant \delta(f,\mathcal{H}_U).\end{aligned}$$

Similarly,

$$|\bar{f}(B)-\bar{f}(AB)|\leqslant\delta(f,\mathcal{H}_V),$$

and hence $|\bar{f}(A)-\bar{f}(B)|\leqslant\delta(f,\mathcal{H}_U)+\delta(f,\mathcal{H}_V)$.

Now we construct the integration. Let $f\in C(G)$. Let U_n be a sequence of open sets such that $\delta(f,\mathcal{H}_{U_n})\longrightarrow 0$ (such a sequence exists by the continuity of f and the compactness of G). Choose, for each n, a U_n-net A_n with minimum cardinality. Then by the Claim, the sequence $\{\bar{f}(A_n)\}$ satisfies the Cauchy convergence criterion and hence it has a limit $L(f)$. The Claim also implies that this limit is independent of the choice of the sets U_n and A_n.

We verify conditions (1)–(4). For (1), note that we can choose a sequence V_n such that simultaneously

$$\delta(\alpha f+\beta g,\mathcal{H}_{V_n})\to 0,\qquad \delta(f,\mathcal{H}_{V_n})\to 0,\qquad \delta(g,\mathcal{H}_{V_n})\to 0.$$

Let A_n be a minimum V_n-net. Then

$$\begin{aligned}L(\alpha f+\beta g)&=\lim_{n\to\infty}\left(\alpha\bar{f}(A_n)+\beta\bar{g}(A_n)\right)\\&=\alpha\lim_{n\to\infty}\bar{f}(A_n)+\beta\lim_{n\to\infty}\bar{g}(A_n)=\alpha L(f)+\beta L(g).\end{aligned}$$

Conditions (2) and (3) are trivial; also, (4) is clear since the whole construction is invariant under translations. □

4. The minimal degree of primitive permutation groups

The *minimal degree* of a permutation group G is the smallest number of points moved by any non-identity element of G. This parameter has been subject to considerable study since the 19th century [see Wielandt (1964) or the forthcoming book by J.O. Dixon and B. Mortimer, *Permutation Groups*].

The minimal degree of the symmetric group is 2 and that of the alternating group is 3. In contrast, there are only finitely many primitive permutation groups with minimal degree μ for $\mu>3$, and none for $\mu=9,25$ or 49. These results were proved by Jordan in 1871 and 1874 (see Jordan 1961). He showed (essentially) that if G is a primitive permutation group of degree n not containing A_n, then

$$\mu\geqslant(1+\mathrm{o}(1))\sqrt{\frac{8n}{\log n}}.$$

Much more is true for doubly transitive groups. Bochert (1889) proved that if G is a doubly transitive permutation group of degree n not containing A_n, then

$$\mu\geqslant\frac{n}{4}-1.$$

The question remains, can Jordan's bound be improved for *uniprimitive* (primitive but not doubly transitive) permutation groups? Consider the line graph $L(K_{v,v})$

of the complete bipartite graph. The automorphism group of this graph is primitive of degree $n = v^2$, and has minimal degree $2v = 2\sqrt{n}$ (consider the automorphism of $L(K_{v,v})$ induced by the transposition of two non-adjacent nodes of $K_{v,v}$).

This example shows that the following result of Babai (1981) is best possible up to the constant.

Theorem 4.1. *If G is a uniprimitive group of degree n then*

$$\mu(G) \geqslant \tfrac{1}{2}\sqrt{n-1}.$$

Below we prove the somewhat weaker inequality $\mu(G) \geqslant \sqrt{(n-1)/6}$. A slight refinement of the proof would yield $\mu(G) \geqslant (1 + o(1))\sqrt{2n/3}$.

Babai proved the theorem by reducing the group-theoretical problem to a combinatorial question on coherent configurations. We shall give this reduction and then a rather simple proof of the result on coherent configurations, incorporating some ideas of Zemlyachenko (see Zemlyachenko et al. 1985).

We need some notions from the theory of coherent configurations (see also chapter 15). A *coherent configuration* $\Xi = (\Omega; R_1, \ldots, R_r)$ is a finite set Ω of vertices and a family R_1, ..., R_r of non-empty binary relations on Ω such that

(i) $\{R_1, \ldots, R_r\}$ forms a partition of $\Omega \times \Omega$;

(ii) the diagonal $\Delta = \{(\omega, \omega) : \omega \in \Omega\}$ is the union of some of the R_i;

(iii) R_i^{-1} (the inverse of R_i) is one of the R_j;

(iv) there is a collection of r^3 integers p_{ij}^k such that for every $\alpha, \beta \in R_k$, one has

$$|\{\gamma\colon\ (\alpha,\gamma) \in R_i,\ \ (\gamma,\beta) \in R_j\}| = p_{ij}^k$$

(independently of the particular choice of α and β).

For $(\alpha, \beta) \in R_i$ we set $i = c(\alpha, \beta)$ and call (α, β) an "edge of color i". We call r the rank of Ξ. The digraphs (Ω, R_i) are the *classes* of Ξ. A coherent configuration is *homogeneous* if $R_1 = \Delta$. It is *primitive* if it is homogeneous and each of the classes (Ω, R_i) $(i \geqslant 2)$ is connected as a digraph (here we do not have to worry which definition of connectedness to choose, since for the classes of a homogeneous coherent configuration connectivity and strong connectivity are equivalent). A primitive coherent configuration is *uniprimitive* if its rank is at least 3.

The classical examples of coherent configurations arise from permutation groups. If G is a permutation group acting on Ω, we let $R_1, \ldots, R_k$ be the orbits of the induced action on $\Omega \times \Omega$, to get a coherent configuration $\Xi(G)$ associated with the permutation group G.

Choosing the indices appropriately, $\Xi(G)$ is homogeneous iff G is transitive. Connected components of any non-diagonal class correspond to blocks of imprimitivity of G (i.e., classes of G-invariant partitions); hence it is easy to see that G is primitive (uniprimitive) iff the associated coherent configuration is primitive (uniprimitive).

The key result of Babai is the following. We say that a vertex ω *distinguishes* vertices α and β if $c(\alpha, \omega) \neq c(\beta, \omega)$.

Theorem 4.2. *Let Ξ be a uniprimitive coherent configuration. Then each pair of vertices α, β is distinguished by at least $\sqrt{(n-1)/6}$ vertices.*

To see that this result implies Theorem 4.1, assume that $\Xi = \Xi(G)$ and let $g \in G$ be a permutation moving only a set M of $\mu(G)$ elements. Let $\alpha \in M$ and $\beta = \alpha^g$. Then for $\omega \notin M$ we have $c(\alpha, \omega) = c(\beta, \omega)$, as g maps (α, ω) onto (β, ω). So α and β are distinguished by at most μ elements. So Theorem 4.2 implies that $\mu \geqslant \sqrt{(n-1)/6}$.

The main idea in proving Theorem 4.1 was to find the right combinatorial relaxation of the group-theoretic notion of minimum degree. As we shall see, the rest of the proof (i.e., the proof of Theorem 4.2) uses only elementary combinatorics.

To prove Theorem 4.2, we need some notation and a series of simple lemmas. For $\omega \in \Omega$, consider the number of edges of color i leaving ω. As Ξ is homogeneous, this number does not depend on ω, and we denote it by d_i. We write $i^{-1} = j$ if $R_i^{-1} = R_j$. In this case $d_i = d_j$. We denote by X_i the digraph (Ω, R_i) and by X_i', the undirected graph $(\Omega, R_i \cup R_i^{-1})$. Let $\operatorname{diam}(i)$ denote the diameter of X_i'.

For two vertices α, β, let $D(\alpha, \beta)$ denote the set of vertices distinguishing α and β. Note that the cardinality of $D(\alpha, \beta)$ depends only on the color of the pair (α, β) (by the definition of coherent configurations); we set $f(i) = |D(\alpha, \beta)|$ if $i = c(\alpha, \beta)$. Clearly $f(i) = f(i^{-1})$. We want to prove that $f(i) \geqslant \sqrt{(n-1)/6}$.

Lemma 4.3. *Let t denote the distance of vertices α and β in the (undirected) graph X_i'. Then $f(i) \geqslant |D(\alpha, \beta)|/t$.*

Proof. Let $\alpha = \alpha_0, \alpha_1, \ldots, \alpha_t = \beta$ be a path of length t in X_i'. Clearly $D(\alpha, \beta) \subseteq \bigcup_{s=1}^{t} D(\alpha_{i-1}, \alpha_i)$. Using $f(i) = f(i^{-1})$ we obtain $|D(\alpha, \beta)| \leqslant t \cdot f(i)$. □

This observation leads to the following important inequality:

Lemma 4.4. *If* $\operatorname{diam}(i) \geqslant 3$ *then* $f(i) \geqslant 2d_i/3$.

Proof. Let α and β be vertices at distance 3 in X_i'. The sets of neighbors (in X_i') of α and β are disjoint, therefore these sets are contained in $D(\alpha, \beta)$. This yields $|D(\alpha, \beta)| \geqslant 2d_i$. So Lemma 4.3 implies that $f(i) \geqslant 2d_i/3$. □

This lemma settles the case when X_i' has diameter at least 3 and d_i is "large". For the case when d_i is "small" we need another simple observation. Let $\Gamma_i(\alpha)$ denote the set of vertices β with $c(\alpha, \beta) = i$.

Lemma 4.5. *If there is an edge with color h from $\Gamma_i(\alpha)$ to $\Gamma_j(\alpha)$ then there are at least* $\max\{d_i, d_j\}$ *such edges.*

Proof. The assumption says that for some vertices β and γ we have $c(\alpha, \beta) = i$, $c(\alpha, \gamma) = j$ and $c(\beta, \gamma) = h$. By the definition of coherent configurations, this implies that for every $\beta \in \Gamma_i(\alpha)$ there exists at least one vertex γ such that $c(\alpha, \gamma) = j$ and $c(\beta, \gamma) = h$. Hence there are at least d_i edges of color h from $\Gamma_i(\alpha)$ to $\Gamma_j(\alpha)$. Applying the argument to h^{-1}, the assertion follows. □

Lemma 4.6. *Any vertex distinguishes the endpoints of at least $n-1$ edges of each color.*

Proof. Consider a vertex ω and a color h, and define a graph W on vertex set $\{1,\dots,r\}$ by joining i to j if there is an edge of color h from $\Gamma_i(\omega)$ to $\Gamma_j(\omega)$. This graph is connected since X_h is connected. Let T be a spanning tree of W. Orient the edges of T away from 1. By Lemma 4.5, every edge ij of T represents at least d_j edges of color h distinguished by ω. Hence the total number of edges of color h distinguished by ω is at least

$$\sum_{ij\in T} d_j = \sum_{j=2}^{r} d_j = n-1. \qquad \square$$

Counting the triples (α,β,ω) with $c(\alpha,\beta)=i$ and $\omega\in D(\alpha,\beta)$ in two ways, and using Lemma 4.6 we obtain

$$nd_i f(i) \geqslant n(n-1),$$

which yields one of our key inequalities:

Lemma 4.7. $f(i)\geqslant (n-1)/d_i$.

Combining with Lemma 4.4, we obtain:

Lemma 4.8. *If* $\operatorname{diam}(i)\geqslant 3$ *then* $f(i)\geqslant\sqrt{2(n-1)/3}$.

What remains is to find a good lower bound on $f(i)$ in the case when $\operatorname{diam}(i)=2$.

Lemma 4.9. *There exists a pair distinguished by at least* $\sqrt{2(n-1)/3}$ *vertices.*

Proof. Suppose that $d_2\leqslant d_3\leqslant\cdots$. If $\operatorname{diam}(2)\geqslant 3$ we are done by Lemma 4.8. If $\operatorname{diam}(2)=2$ then obviously $1+d_2+(d_2-1)d_2\geqslant n$ and hence $d_2\geqslant\sqrt{n-1}$. Further, trivially $d_2\leqslant(n-1)/2$.

Any given vertex distinguishes at least $2d_2(n-1-d_2)\geqslant 2(n-1)(\sqrt{n-1}-1)$ ordered pairs. Hence there must be an ordered pair distinguished by at least $2(\sqrt{n-1}-1)$ vertices, i.e., $f_{\max}\geqslant 2(\sqrt{n-1}-1)>\sqrt{2(n-1)/3}$. $\square$

Proof of Theorem 4.2. Choose a pair α,β of vertices with $|D(\alpha,\beta)|\geqslant\sqrt{2(n-1)/3}$. Let $i\geqslant 2$. If $\operatorname{diam}(i)>2$ then we have $f(i)\geqslant\sqrt{2(n-1)/3}$ by Lemma 4.8. If $\operatorname{diam}(i)=2$ then we have

$$f(i)\geqslant\frac{1}{2}|D(\alpha,\beta)|\geqslant\sqrt{\frac{n-1}{6}}$$

by Lemma 4.3. $\square$

Invoking the classification theorem of finite simple groups, Liebeck and Saxl (1991) obtained a classification of all primitive permutation groups with $\mu\leqslant n/3$.

From this, Theorem 4.1 can be read off. Still, it is good to have a short solution for a classical problem, which helps us understand why the result holds.

The situation is similar for a related classical problem. Using Theorem 4.2 and a simple probabilistic argument, Babai (1981) proved that every uniprimitive permutation group of degree n has a base of size $b \leqslant 4\sqrt{n}\log n$. (A base is a subset Φ of Ω such that the only group element fixing every vertex in Φ is the identity.) This immediately gives an upper bound $n^b \leqslant \exp(4\sqrt{n}\log^2 n)$ on the order of uniprimitive permutation groups, another remarkable result.

This bound has been supplemented (Babai 1982) by an even stronger upper bound of $\exp(\exp(1.08\sqrt{\log n}))$ on the order of doubly transitive permutation groups not containing A_n. Recently Pyber (1993) found a simple combinatorial proof of the bound $n^{c(\log n)^2}$ using Babai's ideas. Once more the classification theorem of finite simple groups gives a stronger bound of $n^{(1+o(1))\log n}$, but perhaps less insight.

5. Algebras with straightening law

In this section, we give a brief glimpse into applications of combinatorial techniques to some questions of algebraic geometry and commutative algebra.

Graded commutative rings (that is, quotients $A = \boldsymbol{k}[x_1,\dots,x_m]/(f_1,\dots,f_m)$, where the f_i are homogeneous polynomials) are a central object of study for commutative algebra. A main motivation for this comes from algebraic geometry, which studies *projective varieties* (that is, solution sets of systems $f_1 = 0,\dots,f_m = 0$ of homogeneous polynomial equations). The geometry of such a variety is encoded in its *coordinate ring* $A = \boldsymbol{k}[x_1,\dots,x_m]/I$, where I is the ideal of all polynomial functions that vanish on the variety, and A can be interpreted as the ring of functions on the variety (see for example Shafarevich 1977, Kunz 1985).

For the study of projective varieties and their coordinate rings, algebraic geometry has a variety of tools available: algebraic, homological, analytic, etc. Combinatorics comes into play in the case of "classical" varieties, such as Grassmann and flag manifolds, determinantal and Schubert varieties, and many more. These varieties and their rings have additional structure: for example, the varieties are very symmetric, and they satify nice smoothness, completeness and regularity conditions. The coordinate rings can be given by generators and relations in a very explicit way, and so it is not too surprising if combinatorial techniques can be applied.

Cohen–Macaulay rings

Let A be a graded commutative ring, that is, the quotient of a polynomial ring modulo a homogeneous ideal,

$$A = \boldsymbol{k}[x_1,\dots,x_m]/I.$$

In particular A has a direct sum decomposition $A = \bigoplus_{k\geqslant 0} A_k$, where A_k is the finite-dimensional $\boldsymbol{k}$-vector space of residue classes of homogeneous polynomials of degree k in A.

In this exposition we will, as an example, consider the Cohen–Macaulay property of A, which expresses a quite subtle regularity property of the associated variety. The usual definition is in homological algebra terms: A is *Cohen–Macaulay* if and only if its *depth* is equal to its *dimension*. Here the dimension $d = \dim(A)$ is the maximal number of algebraically independent elements $\{\theta_1, \ldots, \theta_d\}$ in A, and the depth is the maximal length $p = \operatorname{depth}(A)$ of a *regular sequence:* a sequence $(\theta_1, \ldots, \theta_p)$ such that θ_i is not a zero-divisor in $A/\langle\theta_1, \ldots, \theta_{i-1}\rangle$ for $1 \leqslant i \leqslant p$. This in particular means that $p \leqslant d$: the depth never exceeds the dimension.

There are many reformulations of the Cohen–Macaulay condition, the most explicit perhaps being the following. A is Cohen–Macaulay if and only if it has a *Hironaka decomposition*: for some set $\{\theta_1, \ldots, \theta_d\}$ of algebraically independent, homogeneous elements of A there exist homogeneous elements $\eta_1, \ldots, \eta_t$ such that A has a direct sum decomposition

$$A = \bigoplus_{i=1}^{t} \eta_i \boldsymbol{k}[\theta_1, \ldots, \theta_d].$$

This means that A is a free $\boldsymbol{k}[\theta_1, \ldots, \theta_d]$-module, and the set of *separators* $\{\eta_1, \ldots, \eta_t\}$ forms a module basis for A. The maximal regular sequence $(\theta_1, \ldots, \theta_d)$ is called a *system of parameters* in this case. (It turns out that if A is Cohen–Macaulay, then *any* sequence $(\theta_1, \ldots, \theta_d)$ with $\dim_{\boldsymbol{k}} A/\langle\theta_1, \ldots, \theta_d\rangle < \infty$ can be used as the system of parameters of a Hironaka decomposition.)

To show that the coordinate rings of classical varieties are Cohen–Macaulay, this offers two alternative approaches. One can compute their dimensions and depths – or one can try to construct explicit Hironaka decompositions. The second approach is more far-reaching: a Hironaka decomposition of A carries a lot of extra structure, such that for example the Hilbert function of the ring can be read off. For this we denote the *Hilbert function* of A by

$$\mathrm{H}(A,k) := \dim_{\boldsymbol{k}} A_k,$$

and its *Poincaré series* (or *Hilbert series*) by

$$\operatorname{Poin}(A,t) := \sum_{k\geqslant 0} \mathrm{H}(A,k)t^k,$$

so the Poincaré series of A is the ordinary generating function (see chapter 21) of the Hilbert function.

A basic result (that can for example be derived from the existence of a finite free resolution of A) now states that the Poincaré series is a rational function of the form

$$\operatorname{Poin}(A,t) = \frac{P_A(t)}{(1-t)^d},$$

where $P_A(t)$ is a polynomial in t with integer coefficients, $P_A(0) = 1$ and $P_A(1) \neq 0$, such that the order of the pole of $\operatorname{Poin}(A,t)$ at $t = 1$ is $d = \dim(A)$.

From this in turn we get a basic property of the Hilbert function: for large k, $\mathrm{H}(A,k)$ is a polynomial in k of degree $d-1$, with rational coefficients.

Now, if A is Cohen–Macaulay with $A = \bigoplus_{i=1}^{s} \eta_i \boldsymbol{k}[\theta_1,\ldots,\theta_d]$, then it is easy to read off the Poincaré series

$$\mathrm{Poin}(A,t) = \frac{\sum_{i=1}^{s} t^{\deg \eta_i}}{(1-t)^d}.$$

Note that the numerator polynomial $P_A(t)$ has non-negative coefficients in this Cohen–Macaulay case – this is not true in general. We can also compute the Hilbert function

$$\mathrm{H}(A,k) = \sum_{i=1}^{s} \binom{k - \deg \eta_i + d - 1}{d-1}$$

(with the convention that the summands are zero if $k < \deg \eta_i$), which is a polynomial in k for $k \geqslant \max(\deg \eta_i)$.

Let us mention that more general decompositions allow us to treat quite general classes of rings in a similar way, as is described in Bacławski and Garsia (1981). Computational aspects of this framework are treated in Sturmfels and White (1991).

The straightening law

A common structural feature of many of the important coordinate rings of "classical varieties" (and all those listed above) is that they have the structure of an "algebra with straightening law" over a finite poset – such that the combinatorics of the poset allows us to construct decompositions of the algebra A. A reasonable level of generality is given by the following definition. It describes what is called an *ordinal Hodge algebra* by De Concini et al. (1982), and an *algebra with strongly lexicographic straightening law based on a poset* by Baclawski (1981). See also De Concini and Procesi (1981) and Eisenbud (1980) for introductions, and De Concini and Lakshmibai (1981) for a more general set-up.

Definition 5.1. Let A be a $\boldsymbol{k}$-algebra and $P = \{x_1,\ldots,x_m\}$ a finite poset. Then A is an *algebra with straightening law over P* if A has a presentation of the form $A = \boldsymbol{k}[x_1,\ldots,x_m]/I$ (identifying the elements of P with generators of A) such that

(1) the products of variables $x_{i_1}x_{i_2}\cdots x_{i_k}$ that correspond to multichains $x_{i_1} \leqslant \cdots \leqslant x_{i_k}$ of P, called *standard monomials*, form a $\boldsymbol{k}$-basis of A, and

(2) the *non-standard monomials* can be *straightened*: if for two incomparable elements x_i and x_j of P the monomial x_ix_j is written as a linear combination of standard monomials, then every standard monomial in the expansion contains a variable that is smaller than x_i and a variable that is smaller than x_j.

For every poset $P = \{x_1,\ldots,x_m\}$ we get a canonical algebra with straightening law, called the *Stanley–Reisner ring of P*, as

$$\boldsymbol{k}[P] := \boldsymbol{k}[x_1,\ldots,x_m]/I_P,$$

where I_P is the ideal generated by the non-standard monomials $x_i x_j$, corresponding to incomparable pairs in P. In this ring the non-standard monomials are all zero, so that the straightening law is trivial.

If A is an algebra with straightening law over P, then A and $\boldsymbol{k}[P]$ are isomorphic as $\boldsymbol{k}$-vector spaces. However, $\boldsymbol{k}[P]$ is endowed with a rather trivial multiplication: the product of two standard monomials is either again standard or it is zero (depending on whether the corresponding union of two multichains is again a multichain or not).

Thus the whole algebra structure of $\boldsymbol{k}[P]$ is determined by the combinatorial data of P. This is the case which allows a complete combinatorial analysis. The main result now reduces the general case of an algebra with straightening law to the corresponding Stanley–Reisner ring.

Theorem 5.2 (Bacławski and Garsia 1981; De Concini et al. 1982). *Let A be an algebra with straightening law over P, and $\boldsymbol{k}[P]$ the corresponding Stanley–Reisner ring. Then every Hironaka decomposition $A = \bigoplus_{i=1}^{t} \eta_i \boldsymbol{k}[\theta_1, \ldots, \theta_d]$ of $\boldsymbol{k}[P]$ induces a Hironaka decomposition of A via the canonical vector space isomorphism $A \cong \boldsymbol{k}[P]$. In particular, if the Stanley–Reisner ring $\boldsymbol{k}[P]$ is Cohen–Macaulay, then A is also Cohen–Macaulay.*

Thus in order to prove that an algebra A with straightening law over a poset P is Cohen–Macaulay, it suffices to establish that the "combinatorial" algebra $\boldsymbol{k}[P]$ is Cohen–Macaulay. If parameters θ_i and separators η_j for $\boldsymbol{k}[P]$ have been constructed, then *the same* parameters and separators also yield a Hironaka decomposition of A.

The work by De Concini et al. contains a careful algebraic analysis of the transition from A to $\boldsymbol{k}[P]$. It shows in particular that in fact both algebras have the same dimension, whereas only $\operatorname{depth}(A) \geqslant \operatorname{depth}(\boldsymbol{k}[P])$ holds in general.

Cohen–Macaulay posets

At this stage of the argument, combinatorial methods take over to decide whether $\boldsymbol{k}[P]$ is Cohen–Macaulay, and possibly to construct decompositions.

Definition 5.3. P is a *Cohen–Macaulay poset* (*with respect to* $\boldsymbol{k}$) if $\boldsymbol{k}[P]$ is a Cohen–Macaulay algebra.

Cohen–Macaulay posets were introduced independently by Bacławski and Stanley, aroun' 1975. Subsequent intensive research has produced a wealth of powerful
res'' ia and techniques. For surveys, we refer to Björner et al. (1982) and to
cal Methods" chapter 34 by Björner, where Cohen–Macaulay posets
ome detail (emphasizing the topological approach). Historical de-
acts needed in the following are also discussed there. We will here
cts.

'ery strong conditions on the combinatorics of a poset to allow
aulay property. In fact, P has to be *ranked* (of rank $r(P) =$

$r = \dim k[P]$): this means that every maximal chain $x_1 < x_2 < \cdots < x_r$ in P has the same length $r-1$. Given such a maximal chain, we will write $r(x_i) = i$ for $1 \leqslant i \leqslant r$. This defines the rank $r(x)$ uniquely for every $x \in P$. A *rank selection* of P is a subposet $P_S := \{x \in P\colon r(x) \in S\}$ for some $S \subseteq [r]$. Also, denote by $\hat{P}$ the poset obtained by adding a new maximal element $\hat{1}$ and a new minimal element $\hat{0}$ to P.

Proposition 5.4. *If P is a Cohen–Macaulay poset, then so is every rank selection of every interval of $\hat{P}$.*

Now a strong numerical test has to be satisfied by P (and all its rank selections of intervals). For this, say that the *Möbius function* (see chapter 21, and Stanley 1986) *alternates on P* if for any $x, y \in \hat{P}$ one gets $\mu_{\hat{P}}(x,y) \cdot (-1)^{r(y)-r(x)} \geqslant 0$.

Proposition 5.5. *If P is Cohen–Macaulay, then the Möbius function alternates on P.*

Secondly, there is a complete characterization of Cohen–Macaulay posets in terms of the topology of the simplicial complex $\Delta(P)$ of chains in P, given in section 10.8 of chapter 34. In particular, Cohen–Macaulayness is a topological invariant of $|\Delta(P)|$.

And finally, the techniques of *shellability* and *lexicographical shellability* allow us to explicitly prove Cohen–Macaulayness and to construct Hironaka decompositions for all major classes of Cohen–Macaulay posets. We will use the following formulation of shellability.

Definition 5.6. A poset P is *shellable* if it is ranked and its set $\mathcal{M}$ of maximal chains admits a linear ordering $\mathcal{M} = (C_1, C_2, \ldots, C_t)$ such that every chain C_i contains a unique minimal subset F_i that is not contained in any previous chain C_j $(j < i)$.

The definition immediately implies that for the first chain C_1 one gets $F_1 = \emptyset$, whereas every chain C_i with $i > 1$ satisfies $F_i \neq \emptyset$ and contains at most one "new" vertex. It is not quite trivial to see that the Möbius function alternates on $\hat{P}$. However, the following result implies this.

Theorem 5.7. (Hochster 1972; Stanley 1975; Garsia 1980; Kind and Kleinschmidt 1979). *If P is shellable, then it is Cohen–Macaulay with respect to every field $\mathbf{k}$. In this case we can derive parameters $\theta_i := \sum_{r(x)=i} x$ for $1 \leqslant i \leqslant r$ and separators $\eta_j := \prod_{x \in F_j} x$ for $1 \leqslant j \leqslant t$ from a shelling of P as above.*

This yields, for any shellable poset P, an explicit Hironaka decomposition of $\mathbf{k}[P]$ and thus – with Theorem 5.2 – of any algebra with straightening law over P.

For shellability techniques we again refer to chapter 34 and the references given there. It turns out that very general classes of posets can be shown to be shellable, see section 11.10 in chapter 34. A very powerful result that covers the posets arising from many classical varieties was achieved by Björner and Wachs.

Theorem 5.8 (Björner and Wachs 1982). *The Bruhat order of any finite quotient of a Coxeter group by a parabolic subgroup is (lexicographically) shellable.*

Example: Grassmann varieties

We will illustrate this approach by the most "classical" case of the coordinate rings of Grassmann varieties.

For this let $\boldsymbol{k}$ be a field of characteristic zero, and consider the pth exterior product of $\boldsymbol{k}^n$, denoted $\Lambda_p\boldsymbol{k}^n$. It is a $\boldsymbol{k}$-vector space of dimension $\binom{n}{p}$, with an explicit basis given by $\{\boldsymbol{e}_{i_1} \wedge \cdots \wedge \boldsymbol{e}_{i_p}\colon 1 \leqslant i_1 < \cdots < i_p \leqslant n\}$. The corresponding coordinate functions for $\Lambda_p\boldsymbol{k}^n$ are denoted by $[i_1 \cdots i_p]$. Thus a general (antisymmetric) tensor in $\Lambda_p\boldsymbol{k}^n$ has the form $\omega = \sum_{1\leqslant i_1<\cdots<i_p\leqslant n}[i_1 \ldots i_p]\ \boldsymbol{e}_{i_1} \wedge \cdots \wedge \boldsymbol{e}_{i_p}$.

An *extensor* is a non-zero, decomposable tensor in $\Lambda_p\boldsymbol{k}^n$, that is, a tensor of the form $\omega = \boldsymbol{v}_1 \wedge \cdots \wedge \boldsymbol{v}_p$ for p linearly independent vectors $\boldsymbol{v}_i \in \boldsymbol{k}^n$. Two p-tuples $(\boldsymbol{v}_1,\ldots,\boldsymbol{v}_p)$ and $(\boldsymbol{v}'_1,\ldots,\boldsymbol{v}'_p)$ determine up to a scalar the same extensor (that is, $\boldsymbol{v}_1 \wedge \cdots \wedge \boldsymbol{v}_p = c\cdot\boldsymbol{v}'_1 \wedge \cdots \wedge \boldsymbol{v}'_p$ for some $c \neq 0$) if and only if they span the same p-dimensional subspace of $\boldsymbol{k}^n$.

In coordinates, for $\boldsymbol{v}_i = \sum_{j=1}^n v_{ij}\boldsymbol{e}_j$, one gets

$$\boldsymbol{v}_1 \wedge \cdots \wedge \boldsymbol{v}_p = \sum_{1\leqslant i_1<\cdots<i_p\leqslant n} [i_1 \cdots i_p]\boldsymbol{e}_{i_1} \wedge \cdots \wedge \boldsymbol{e}_{i_p},$$

$$\text{where } [i_1 \cdots i_p] = \begin{vmatrix} v_{1i_1} & v_{1i_2} & \cdots & v_{1i_p} \\ v_{2i_1} & v_{2i_2} & \cdots & v_{2i_p} \\ \vdots & \vdots & \ddots & \vdots \\ v_{pi_1} & v_{pi_2} & \cdots & v_{pi_p} \end{vmatrix}.$$

The vector of $\binom{n}{p}$ coordinates $([i_1 \cdots i_p]\colon 1 \leqslant i_1 < \cdots < i_p \leqslant n)$ is called a vector of *Plücker coordinates* for $V := \operatorname{span}(\boldsymbol{v}_1,\ldots,\boldsymbol{v}_p)$.

Passing to the projective space of $\Lambda_p\boldsymbol{k}^n$, we find that the set

$$G_{p,n} := \{[\boldsymbol{v}_1 \wedge \cdots \wedge \boldsymbol{v}_p] \in P(\Lambda_p\boldsymbol{k}^n)\}$$

is in natural bijection to the set of p-subspaces of $\boldsymbol{k}^n$. As the image of a homogeneous polynomial map (which sends the matrix (v_{ij}) to the vector $([i_1 \cdots i_p])$ of all its maximal minors) this $G_{p,n}$ is an irreducible projective variety, called the *Grassmann variety* of p-subspaces in $\boldsymbol{k}^n$. This variety turns out to have dimension $p(n-p)$, in an ambient space of dimension $\binom{n}{p} - 1$. Note that $G_{1,n} = P(\boldsymbol{k}^n)$ is projective space.

The coordinate ring of the projective variety $G_{p,n}$ is

$$A(G_{p,n}) = \boldsymbol{k}\big[[i_1 \cdots i_p] : 1 \leqslant i_1 < \cdots < i_p \leqslant n\big]/I_{p,n},$$

where $I_{p,n}$ is the ideal of relations between the $(p \times p)$-minors of a generic $(p \times n)$-matrix, that is, between the variables $[i_1 \cdots i_p]$ if they are given by

$$[i_1 \cdots i_p] := \begin{vmatrix} v_{1i_1} & \cdots & v_{1i_p} \\ \vdots & \ddots & \vdots \\ v_{pi_1} & \cdots & v_{pi_p} \end{vmatrix}.$$

Monomials of degree t in $A(G_{p,n})$ are customarily denoted by $(t \times p)$-arrays, called tableaux, whose rows denote the names of the variables. So $[i_1 \cdots i_p] \cdot [j_1 \cdots j_p]$ will be denoted as $\left[\begin{smallmatrix} i_1 & \cdots & i_p \\ j_1 & \cdots & j_p \end{smallmatrix}\right]$, etc.

The basic theorem now is that the ideal $I_{p,n}$ is generated by the *straightening syzygies*:

$$\begin{bmatrix} i_1 \cdots i_p \\ j_1 \cdots j_p \end{bmatrix} = -\sum_{\sigma} \operatorname{sign}(\sigma) \begin{bmatrix} i_1 \cdots i_{l-1} & \sigma(i_l) & \cdots \sigma(i_p) \\ \sigma(j_1) \cdots & \sigma(j_l) & j_{l+1} \cdots j_p \end{bmatrix}, \qquad (*)$$

where l is an arbitrary fixed position ($1 \leqslant l \leqslant p$), and the sum is over all shuffles σ of $(j_1 \cdots j_l | i_l \cdots i_p)$ except for the identity, that is, over all permutations $\sigma \neq \mathrm{id}$ of the (multi)set $\{j_1, \ldots, j_l, i_l, \ldots, i_p\}$ that satisfy $\sigma(j_1) < \cdots < \sigma(j_l)$ and $\sigma(i_l) < \cdots < \sigma(i_p)$. Here we define $[i_{\sigma(1)} \cdots i_{\sigma(p)}] := \operatorname{sign}(\sigma)[i_1 \cdots i_p]$ for p-tuples that are not increasing, and $[i_1 \cdots i_p] = 0$ whenever two entries i_l are equal.

[In fact, the ideals $I_{p,n}$ are already generated by the *Grassmann–Plücker relations* that are obtained from $(*)$ in the special case $l = 1$. However, these are not sufficient for straightening – see Sturmfels and White (1989).]

We want to conclude that $A(G_{p,n})$ is an algebra with straightening law. For this we define a partial order on the set of variables, by putting $[i_1 \cdots i_p] \leqslant [j_1 \cdots j_p]$ whenever $i_1 \leqslant j_1, \ldots, i_p \leqslant j_p$. This makes

$$\Lambda_{p,n} := \left(\{[i_1 \cdots i_p] : 1 \leqslant i_1 < \cdots < i_p \leqslant n\}, \leqslant\right)$$

into a partial order – in fact, $\Lambda_{p,n}$ turns out to be a distributive lattice. In particular, $\Lambda_{p,n}$ is ranked: its rank function is given by $r([i_1 \cdots i_p]) = 1 + (i_1 - 1) + \cdots + (i_p - p)$, and $\Lambda_{p,n}$ has length $r([n-p+1, \ldots, n]) = 1 + p(n-p)$.

Furthermore, $\Lambda_{p,n}$ is shellable: shellings for distributive lattices are easily obtained by a lexicographic technique; also – and this is the argument that generalizes for other classical varieties – $\Lambda_{p,n}$ is the quotient of (the Bruhat order of) the Coxeter group $\mathcal{S}_n$ by its maximal parabolic subgroup $\mathcal{S}_p \times \mathcal{S}_{n-p}$ and hence shellable as a special case of the Björner–Wachs theorem 5.8.

Now reconsider the tableau corresponding to a monomial in $A(G_{p,n})$. The rows of such tableaux are variable names and thus strictly increasing, and a tableau is *standard* (that is, corresponds to a standard monomial) exactly if the columns are all non-decreasing (if read from top to bottom).

With the facts we have collected now, it is not hard to verify that $A(G_{p,n})$ is an algebra with straightening law over $\Lambda_{p,n}$: in fact the Grassmann–Plücker relations $(*)$, iterated, show that the standard monomials span $A(G_{p,n})$, and they also provide the straightening law for $A(G_{p,n})$: for this consider any monomial of degree 2 that is not standard, that is, a tableau $\left[\begin{smallmatrix} i_1 & \cdots & i_p \\ j_1 & \cdots & j_p \end{smallmatrix}\right]$ for which $i_l > j_l$ for some l. In this situation one has $j_1 < \cdots < j_l < i_l < \cdots < i_p$, and the formula $(*)$ expresses the tableau as a linear combination of tableaux whose top rows are all smaller than $[i_1 \cdots i_p]$. Iterating this procedure, we get an expression of $\left[\begin{smallmatrix} i_1 & \cdots & i_p \\ j_1 & \cdots & j_p \end{smallmatrix}\right]$ as a linear combination of standard tableaux whose top rows are smaller than $[i_1 \cdots i_p]$. (This shows that the standard monomials span $A(G_{p,n})$.) But we could have applied the same procedure to $\left[\begin{smallmatrix} j_1 & \cdots & j_p \\ i_1 & \cdots & i_p \end{smallmatrix}\right]$ as well, to get the *same* expression again (because the standard monomials form a basis!). Thus the top rows of all the standard tableaux in the expansion of $\left[\begin{smallmatrix} i_1 & \cdots & i_p \\ j_1 & \cdots & j_p \end{smallmatrix}\right]$ are also smaller than $[j_1 \cdots j_p]$, which shows property (2) of an algebra with straightening law.

Thus $G_{p,n}$ is a Cohen–Macaulay variety, much of whose structure is (in a subtle way) controlled by the poset $\Lambda_{p,n}$.

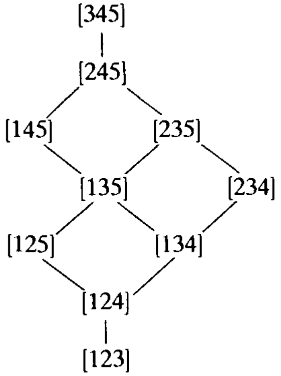
For an explicit example, consider the Grassmann variety $G_{3,5}$. Its coordinate ring is

$$A(G_{3,5}) = \boldsymbol{k}\big[[123],[124],[125],[134],[135],[145],[234],[235],[245],[345]\big]/I_{3,5},$$

where the ideal $I_{3,5}$ is generated by Grassmann–Plücker relations like

$$\begin{bmatrix} 145 \\ 234 \end{bmatrix} - \begin{bmatrix} 134 \\ 245 \end{bmatrix} + \begin{bmatrix} 124 \\ 345 \end{bmatrix} = 0.$$

Now $A(G_{3,5})$ is an algebra with straightening law over the poset $\Lambda_{3,5}$:



It is easy to get a (lexicographic) shelling: take

$$\begin{aligned}
C_1 &= ([123] < [124] < [125] < [135] < [145] < [245] < [345]) \\
C_2 &= ([123] < [124] < [125] < [135] < [235] < [245] < [345]) \\
C_3 &= ([123] < [124] < [134] < [135] < [145] < [245] < [345]) \\
C_4 &= ([123] < [124] < [134] < [135] < [235] < [245] < [345]) \\
C_5 &= ([123] < [124] < [134] < [234] < [235] < [245] < [345])
\end{aligned}$$

This yields parameters $\eta_1 = 1$, $\eta_2 = [235]$, $\eta_3 = [134]$, $\eta_4 = [134][235]$, $\eta_5 = [234]$, and separators $\theta_1 = [123]$, $\theta_2 = [124]$, $\theta_3 = [125] + [134]$, $\theta_4 = [135] + [234]$, $\theta_5 =$

[145] + [235], $\theta_6 = [245]$ and $\theta_7 = [345]$, which describes a Hironaka decomposition of the Cohen–Macaulay algebra $A(G_{3,5})$, with $d = 7$ and $t = 5$:

$$\begin{aligned} A(G_{3,5}) = &\boldsymbol{k}\big[[123],[124],[125]+[134],[135]+[234],[145]+[235],[245],[345]\big] \\ &\oplus [235]\boldsymbol{k}\big[[123],[124],[125]+[134],[135]+[234],[145]+[235],[245],[345]\big] \\ &\oplus [134]\boldsymbol{k}\big[[123],[124],[125]+[134],[135]+[234],[145]+[235],[245],[345]\big] \\ &\oplus [134][235]\boldsymbol{k}\big[[123],[124],[125]+[134],[135]+[234],[145]+[235],[245],[345]\big] \\ &\oplus [234]\boldsymbol{k}\big[[123],[124],[125]+[134],[135]+[234],[145]+[235],[245],[345]\big] \end{aligned}$$

From this we can read off the Poincaré series as

$$\mathrm{Poin}(A(G_{3,5}),t) = \frac{1+3t+t^2}{(1-t)^7},$$

and the Hilbert function as

$$\mathrm{H}(A(G_{3,5}),k) = \binom{k+6}{6} + 3\binom{k+5}{6} + \binom{k+4}{6}.$$

6. Complex hyperplane arrangements

Combinatorial properties of arrangements of hyperplanes in real linear space have been studied in geometry for quite a while. Several properties of such arrangements are described in chapter 17. A more recent development is the application of combinatorial techniques to the theory of complex hyperplane arrangements.

In the course of investigation it has turned out that deep structural properties of arrangements are controlled by their matroids, and combinatorial methods allow us not only to compute important invariants (such as the cohomology algebra of the complement, and the homotopy type of the link) but also identify correctly extreme cases with special structure (supersolvable and 3-generic arrangements).

In the following sketch we focus on the links and connections between *combinatorial* data (hyperplane arrangements as linearly represented matroids), *algebraic structure* (hyperplane arrangements as singular varieties) and *topological aspects* (complements of arrangements as complex manifolds).

We refer to the lecture notes by Orlik (1989) and the book by Orlik and Terao (1992) on arrangements for further details, expositions and extensive bibliographies.

For this section we restrict our attention to the following scenario (although greater generality is possible and natural for many aspects).

Definition 6.1. A *hyperplane arrangement* is a finite set $X = \{H_1, \ldots, H_m\}$ of hyperplanes (vector subspaces of codimension 1) in a complex vector space $V = \mathbb{C}^n$, with $\bigcap_{i=1}^m H_i = \{0\}$.

This definition describes *complex* arrangements (it is also interesting to study arrangements over other fields, the case of arrangements in real vector spaces being the best studied one). The arrangements considered here are *linear*, because all hyperplanes are required to contain the origin, and they are *essential*: the last condition requires that the *dimension* $n = \dim(V)$ coincides with the rank $r = \operatorname{codim}_V(\bigcap_{i=1}^{m} H_i) = n - \dim(H_1 \cap \cdots \cap H_m)$ of the arrangement X. These last two conditions (linear and essential) are not very restrictive – usually questions are easily reduced to this case. There are, in fact, standard constructions that transform from *affine* or *projective* arrangements to linear ones, and every arrangement has a canonical essential arrangement associated with it.

Choosing coordinates $x_1, \ldots, x_n \in V^*$ for V, we write $S := \mathbb{C}[x_1, \ldots, x_n]$ for the ring of polynomial functions on V. For every hyperplane $H \in X$ we can then choose a linear form $\ell_H \in V^*$ whose kernel is H, which may also be seen as a linear function $\ell_H \in S$ that defines $H = \{x \in V: \ell_H(x) = 0\}$.

The product $Q := \prod_{H \in X} \ell_H \in S$ is thus interpreted as a *defining equation* for the arrangement X: it defines the hypersurface obtained by the union of the hyperplanes in X. Note that both Q and the ℓ_H are unique up to a non-zero complex factor.

We will in the following sketch structures associated with hyperplane arrangements by various branches of mathematics. The common questions will be:

- How much of the structure of a hyperplane arrangement is encoded in and determined by its combinatorics?
- Do the combinatorial data allow simple construction or computation of the relevant information?
- Do the combinatorial invariants identify interesting special cases with additional structure?

In the light of these questions, the following three sections will first consider the combinatorics associated with an arrangement, then the topological properties, and finally algebraic structures.

The matroid of an arrangement

The *combinatorial structure* associated with a hyperplane arrangement is that of a *matroid* (cf. chapter 9). [This contrasts the case of real arrangements, where much more information is encoded in the associated *oriented matroid* (cf. chapter 9, section 15, chapter 34, section 7, and Björner et al. 1993).]

Definition 6.2. The *matroid* $M = M_X$ of the arrangement X is given by linear independence on the set $\{\ell_H: H \in X\}$ of vectors in V^*. A property of arrangements is *combinatorial* if it is determined by the matroids of the arrangements alone. M is a matroid of rank $r(M) = n$ on a ground set of $m = |X|$ elements, which we identify with X. Important invariants for the following are its *lattice of flats*, which is isomorphic to the intersection lattice

$$L = \left\{ \bigcap Y : Y \subseteq X \right\}$$

(ordered by reverse inclusion), its *characteristic polynomial*

$$\chi(t) = \sum_{x \in L} \mu(\hat{0}, x) t^{n-r(x)} = \sum_{i=0}^{r} w_i (-1)^i t^{n-i},$$

and the broken circuit complex $\mathrm{BC}(M) \subseteq 2^{\{1,2,\ldots,m\}}$, a simplicial complex with exactly w_i faces of cardinality i, which implies $w_i > 0$ for $0 \leqslant i \leqslant n$.

We will now introduce three classes of hyperplane arrangements that are distinguished not only by combinatorial conditions, but also by algebraic and topological ones, as we will soon see.

- A hyperplane arrangement is *supersolvable* if its matroid is supersolvable [in the sense of Stanley (1972), that is, if its intersection lattice contains a maximal chain of modular elements].
  A key fact is the factorization $\chi(t) = \prod_{i=1}^{n}(t - e_i)$ for positive integers e_i in this case. This can be explained by a factorization of the broken circuit complex which characterizes supersolvability, according to Björner and Ziegler (1991).
  Supersolvable arrangements form a very interesting class of highly structured arrangements. For example, if G is a graph on $\{1, \ldots, n\}$, then the arrangement of hyperplanes $x_i = x_j$ (for $ij \in E(G)$) is supersolvable if and only if the graph is chordal.
- An arrangement is *generic* if arbitrary small perturbations do not change the combinatorial structure. This is equivalent to the property that the matroid M is *uniform*, such that $M \cong U_{n,m}$.
  A weaker condition is that no hyperplane of the arrangement contains the intersection of two other hyperplanes. Equivalently, we may require that M contains no 3-circuits, which defines 3-*generic* arrangements.
  The 3-generic arrangements are a large class of arrangements with very little (combinatorial) structure.
- A third class of "special" (although not combinatorially defined) complex arrangements arises by complexifying simplicial real arrangements (that is, by field extension for arrangements that subdivide $\mathbb{R}^n$ into simplicial cones). Such arrangements arise, for example, from the actions of finite Coxeter groups (groups generated by reflections), when one considers the arrangement of all hyperplanes of reflections in the group.
  Complexified real arrangements can be treated in terms of the combinatorial data given by the real hyperplane arrangements. This is the reason why they are much better understood than general complex ones.

Topology of the complement

Consider the hyperplane arrangement X as a complex hypersurface (of real codimension 2!) in V. Then $T := V \backslash X$, called the *complement* of the arrangement, is a connected open complex manifold with interesting cohomology and homotopy properties. Closely related to this one studies the *link* $D := X \cap S^{2n-1}$, the intersection of the arrangement with the unit sphere in $\mathbb{C}^n$.

Note that for the real analogue (a linear arrangement in $\mathbb{R}^n$), this complement is a union of disjoint open convex cones, and so its topological structure is entirely determined by the number of its components, which is described by Zaslavsky's theorem (see chapter 17). The link has the homotopy type of a wedge of $(n-2)$-spheres. In contrast to this, no complete description of (say, the homotopy type of) the space T is known for the complex case. However, one can describe the homotopy type of the link in this case, see below.

In the following, we will present a complete combinatorial construction of the cohomology algebra of T, and some interesting partial results for the homotopy structure.

The construction for the cohomology algebra can be sketched as follows. Let E be a free $\mathbb{Z}$-module with basis $\{e_1,\dots,e_m\}$ in bijection with X. Let AE be the exterior algebra over E, with basis $\{e_K\colon K\subseteq[n]\}$. Denote by I the ideal of AE which is generated by all the elements of the form $\partial(e_C)$, where C is a circuit of M. Here the boundary map ∂ is defined by linear extension of

$$\partial(e_K)=\sum_{j=1}^{p}(-1)^{j-1}e_{K-\{i_j\}},\quad \text{for } K=\{i_1,i_2,\dots,i_p\}_<.$$

Now define the *Orlik–Solomon algebra* of X as $A(M):=AE/I$. This algebra is combinatorial, since it is constructed from matroid data only. It provides a model for the cohomology algebra of T which is combinatorial in the sense of Definition 6.2.

Theorem 6.3 (Orlik and Solomon 1980). *Let X be a complex arrangement, M its matroid and T its complement. Then the cohomology algebra of T (with integer coefficients) is isomorphic as a graded $\mathbb{Z}$-algebra to the Orlik–Solomon algebra $A(M)$:*

$$H^*(T,\mathbb{Z})\cong A(M).$$

This fundamental theorem allows a complete analysis of the cohomology of T with well-developed combinatorial tools. So one finds that the broken circuit complex of M induces a basis of $A(L)$ and hence of the cohomology algebra $H^*(T,\mathbb{Z})$. This follows from Björner (1982) and Orlik and Solomon (1980) and was rediscovered by Jambu and Terao (1989). In particular this proves that the Betti numbers of T are given by

$$\beta^i(T)=\operatorname{rank}H^i(T,\mathbb{Z})=w_i,$$

so that the Poincaré polynomial of $H^*(T,\mathbb{Z})$ is $t^n\chi(-1/t)$ (Orlik and Solomon 1980). It can also be used for an elementary proof of Theorem 6.3, see Björner and Ziegler (1992).

With observations of this type there is a complete analysis of the cohomology algebra of T with matroid theory tools. Similarly, the following result shows that the homotopy type of the link D is combinatorial.

Theorem 6.4 (Björner and Ziegler 1992, Ziegler and Živaljević 1993). *The link $D = X \cap S^{2n-1}$ of every complex arrangement has the homotopy type of a wedge of spheres, put together from w_i spheres of dimension $2n-2-i$, for $i > 0$:*

$$D \simeq \bigvee_{A \in \mathrm{BC}(M)\setminus\emptyset} S^{2n-2-|A|}.$$

Here for $n = 2$ the wedge is formed to produce a disjoint union of circles. Otherwise, the link is connected and the homotopy type of the wedge does not depend on the choice of wedge points.

The approach of Björner and Ziegler (1992) produces explicit spheres S_A in the link D that induce the homotopy equivalence. The "diagram method" of Ziegler and Živaljević (1993) works in a much more general situation. It produces homotopy formulas for links of arrangements of arbitrary real subspaces, and the above is just a special application of their method.

Using Alexander duality (see Spanier 1966), one gets from Theorem 6.4 the *linear structure* of the cohomology algebra of T. However, the *multiplicative structure* described by Theorem 6.3 cannot be derived from the homotopy type of the link: it encodes subtle details about the complex structure of the arrangement, see Ziegler (1993).

Also, via Spanier–Whitehead duality one gets from Theorem 6.4 that the complement T has the stable homotopy type of a wedge of spheres (i.e., after a sufficient number of suspensions it is homotopy equivalent to a wedge of spheres). However, a complete description of the homotopy type of T seems to be out of reach. In fact, the following basic problems are open, see Falk and Randell (1985), Salvetti (1987).

- Is the homotopy type of a complex arrangement combinatorial? That is, do arrangements with isomorphic matroids always have homotopy-equivalent complements?
- In particular, is the fundamental group $\pi_1(T)$ combinatorial? That is, can one give a presentation of $\pi_1(T)$ from the knowledge of M alone?

Without a positive solution to this problem, more data than just the matroid are necessary to construct the homotopy type of an arrangement. For example, in the case of complexified real arangements, one can construct T *up to homeomorphism* using the additional data given by the oriented matroid of the real arrangement. This was shown by Björner and Ziegler (1992), extending an earlier similar result of Salvetti (1987) for the homotopy type.

In the general complex case, several non-trivial results have been given, and there is very active research going on. See, for example, Arvola (1992) and Falk (1993) for recent progress. Much of the attention centers on the first homotopy group $\pi_1(T)$, together with the question under which conditions T is a $K(\pi, 1)$ space, that is, the higher homotopy groups $\pi_i(T)$ $(i > 1)$ vanish.

Crucial parameters are given by the *lower central series* of $\pi_1(T)$, defined by $c_k := \mathrm{rank}(G_k/G_{k+1})$, where $G_1 = \pi_1(T)$, and $G_{k+1} = [G_k, G_1]$ is the subgroup of G_1 generated by the commutators $ghg^{-1}h^{-1}$ for elements $g \in G_k$ and $h \in G_1$.

Theorem 6.5. *Let T be the complement of a complex arrangement X.*

(1) (Hattori 1975) *If X is generic and $|X| > n \geqslant 2$, then T is not a $K(\pi, 1)$ space: in this case $\pi_i(T) = 0$ for $1 < i < n$, and $\pi_n(T)$ is free Abelian on an infinite number of generators.*

(2) (Deligne 1972) *If X is a complexified simplicial arrangement, then T is a $K(\pi, 1)$ space.*

(3) (Terao 1986, Falk and Randell 1985) *If X is supersolvable, then T is a $K(\pi, 1)$ space, and its lower central series is given by $\prod_{j=1}^{\infty}(1-t^j)^{c_j} = \prod_{k=1}^{n}(1-e_k t) = t^n \chi(1/t)$.*

The combinatorics underlying Deligne's Theorem 6.5(2) was clarified recently by Paris (1993a,b), see also Cordovil (1994) and Salvetti (1993).

It is safe to say that the homotopy groups of T are in general extremely complicated objects. We just mention that the formula in part (3) of this theorem fails even for very nice complexified simplicial arrangements.

The homotopy structure of T is, nevertheless, a very promising field of research, and combinatorial approaches and methods should be helpful to attack some basic open problems, the most striking ones being the following "$K(\pi, 1)$-problems" (of which the last is a special case of the problem mentioned before Theorem 6.5):

- Does Hattori's result (Theorem 6.5(1)) generalize to 3-generic arrangements? Or can T be a $K(\pi, 1)$ for some 3-generic arrangement?
- Describe necessary and sufficient conditions for T to be a $K(\pi, 1)$.
- Does the matroid alone determine whether the complement of a complex arrangement is a $K(\pi, 1)$?

The module of logarithmic vector fields

An *algebraic* structure of interest here is the S-module of algebraic vector fields that are tangent to the hyperplanes. Saito's (1980) investigations in singularity theory first suggested the study of these modules of *logarithmic vector fields* and, dually, of *logarithmic differential forms* at a hypersurface singularity. Specialization to the case of hyperplane arrangements led Terao (1980) to his fascinating theory of *free* hyperplane arrangements.

The following module captures a lot of structure of the arrangement. Its control by combinatorial data is quite strong, but not straightforward.

Definition 6.6. Let X be a complex arrangement defined by $Q = \prod_{H \in X} \ell_H \in S$. The *$S$-module of logarithmic vector fields* at X is the set of derivations

$$\mathrm{Der}(X) = \left\{ \theta = \sum_{i=1}^{n} p_i \frac{\partial}{\partial x_i} : Q | \theta(Q) \right\}$$

with the obvious S-module structure.

There is also a very simple, geometric description: $\mathrm{Der}(X)$ *is isomorphic to the module of all polynomial maps $p : \mathbb{C}^n \longrightarrow \mathbb{C}^n$ that map every hyperplane of X into itself.*

For this, p is considered as an n-tuple $p = (p_1, \ldots, p_n) \in S^n$ of n-variable polynomials; the corresponding element in $\mathrm{Der}(X)$ is $\theta_p = \sum_{i=1}^n p_i(\partial/\partial x_i)$. To see that the set of such maps p is an S-module, one has to show that S-linear combinations have the same form.

The following "basis criterion" of Saito (1980) identifies the case when $\mathrm{Der}(X)$ is a free module (that is, has a basis).

Lemma 6.7. *The logarithmic vector fields* $\theta_1, \ldots, \theta_n \in \mathrm{Der}(X)$ *with* $\theta_i = \sum_{j=1}^n p_{ij}(\partial/\partial x_j)$ *form a basis of* $\mathrm{Der}(X)$ *if and only if* $\det(p_{ij}) = cQ$ *for a non-zero constant* $c \in \mathbb{C}$.

This situation arises in many important examples and provides an algebraic criterion for "strong combinatorial structure" in an arrangement.

Definition 6.8 (Terao 1980). The arrangement X is *free* if $\mathrm{Der}(X)$ is a free S-module.

$\mathrm{Der}(X)$ is an S-module of rank n (every maximal S-linearly independent subset has cardinality S), such that the basis criterion above characterizes (bases of) free arrangements.

At this point we observe that it pays off to also study the dual module $\Omega^1(X) = \mathrm{Hom}_S(\mathrm{Der}(X), S)$ of *logarithmic* 1*-forms at* X, because it has additional structure. $\Omega^1(X)$ is the S-module of differential 1-forms $\omega = (1/Q)\sum_{i=1}^n q^i \mathrm{d}x_i$ with $q^i \in S$, such that $\mathrm{d}\omega$, like ω, has at most a first order pole at every hyperplane $H \in X$. This is equivalent to requiring that the restrictions $(Q\omega)|_H$ vanish. The differential forms $\mathrm{d}\ell_H/\ell_H$ for $H \in X$ are some obvious elements of $\Omega^1(X)$.

There is a non-degenerate pairing between $\mathrm{Der}(X)$ and $\Omega^1(X)$. In particular $\Omega^1(X)$ is free if and only if $\mathrm{Der}(X)$ is free, which offers a second approach to free arrangements.

Both $\mathrm{Der}(X)$ and $\Omega^1(X)$ have the natural structure of a graded module, where $\theta \in \mathrm{Der}(X)$ is homogeneous of degree e if $\theta = \sum_{i=1}^n p_i(\partial/\partial x_i)$, where all the p_i are either 0 or homogeneous of degree e. Similarly, a logarithmic differential form $\omega = (1/Q)\sum_{i=1}^n q^i \mathrm{d}x_i$ has degree $e - m$ if all the polynomials q^i are either zero or homogeneous of degree e: here $m = |X|$ is the degree of Q.

So for every $H \in X$, $\mathrm{d}\ell_H/\ell_H$ is a form of degree -1, whereas $(1/\ell_1)(\mathrm{d}\ell_2/\ell_2 - \mathrm{d}\ell_3/\ell_3)$ is a form of degree -2 in $\Omega^1(X)$ if and only if the linear forms ℓ_1, ℓ_2, and ℓ_3 are dependent, so the corresponding hyperplanes satisfy $H_1 \supseteq (H_2 \cap H_3)$, that is, they form a 3-circuit in M_X.

Now if $\mathrm{Der}(X)$ is free, then it has a basis $\{\theta_1, \ldots, \theta_n\}$ consisting of homogeneous vector fields $\theta_i = \sum_{j=1}^n p_{ij}(\partial/\partial x_j)$ of respective degrees e_i. Similarly, $\Omega^1(X)$ then has a homogeneous basis $\{\omega_1, \ldots, \omega_n\}$ with degrees $\deg(\omega_j) = -e_j$, which is explicitly given as $\omega_j = \sum_{i=1}^n q^{ij} \mathrm{d}x_i$, where the matrices (q^{ij}) and (p_{ij}) are inverses of each other.

Terao's (1981) remarkable "Factorization theorem" now implies that the degrees of these homogeneous basis elements – if X is free – can be computed combinatorially.

Theorem 6.9 (Terao 1981). *If X is a free arrangement, then the characteristic polynomial of M_X factors as*

$$\chi(t) = \prod_{i=1}^{n} (t - e_i),$$

where the e_i are the degrees of the vector fields in a homogeneous basis of $\mathrm{Der}(X)$.

This result in particular allows us to disprove freeness for large classes of examples. It also suggests classes of arrangements that "ought to be free" – because they meet the strong combinatorial criterion that $\chi(t)$ factors over $\mathbb{Z}$.

The original proof of this result was very difficult. New ideas by Solomon and Terao (1987) leading to Rose and Terao's (1990) concept of a "Möbius sum with Hilbert coefficients" put this into a simpler and broader framework.

Theorem 6.9 is part of a bulk of evidence for the following conjecture that has motivated a lot of research on free arrangements.

Conjecture 6.10 (Terao's conjecture). Freeness is a combinatorial property.

Extensive work has gone into this conjecture, without final success, yet. Special cases can be settled: for example, arrangements with a graphic matroid are free if and only if they are supersolvable (Stanley, see Edelman and Reiner 1994), and Terao's conjecture is true if the matroid is binary (Ziegler 1990). Edelman and Reiner (1994) characterized freeness for a larger class of arrangements (described in terms of graphs), which includes both free and nonfree arrangements. Their analysis also provides classes of counterexamples to "Orlik's conjecture": it is *not* true that the restriction of a free arrangement to one of its hyperplanes always a free arrangement; see Edelman and Reiner (1993).

Furthermore, ten years of research have clarified the structure of the Saito–Terao modules quite a bit, and in this course the importance and the scope of the underlying combinatorial structure have become more apparent – see also Ziegler (1990) and Yuzvinsky (1993). It has also led to algebraic characterizations for the two combinatorial classes of arrangements discussed above.

Theorem 6.11 (Stanley 1984; Jambu and Terao 1984; Ziegler 1989). *Let X be a complex arrangement.*

(1) *X is supersolvable if and only if it is free with a basis $\{\theta_1, \ldots, \theta_n\}$ of* $\mathrm{Der}(X)$ *that (in suitable coordinates) has an upper-triangular coefficient matrix: $\theta_i = \sum_{j=i}^{n} p_{ij}(\partial/\partial x_j)$ for $1 \leqslant i \leqslant n$.*

(2) *X is* 3-generic *if and only if the module $\Omega^1(X)$ does not contain forms of degree -2, that is, if it is generated by $\{\mathrm{d}\ell_H/\ell_H\colon H \in X\}$.*

Thus a 3-generic arrangement (and, in particular, a generic arrangement in dimension $n \geqslant 3$) can never be free unless $n = m$.

The role of the complexified simplicial arrangements is less clear in this context. Terao (1980) has shown, analyzing the $n = 3$ case, that many, but not all such

arrangements are free. This was motivated by the
a $K(\pi, 1)$ space for every free arrangement X. This w
Reiner in 1994. The converse was disproved by Terao vers
and Deligne's theorem 6.5(2).

Thus it is still an open problem to directly relate the a
rangements with their homotopy types. A lot of work wil
elaborating on the interpretation and relevance of combinato
plex structure of hypersurface singularities. Combinatorial, di
many of the algebraic and topological properties. To understanc
also in much more general settings than the model of this section,
a deeper understanding of the connections between singularity theo opology, reflection groups, and a zoo of other topics that we have not e n touched upon here.

7. Knots and the Tutte polynomial

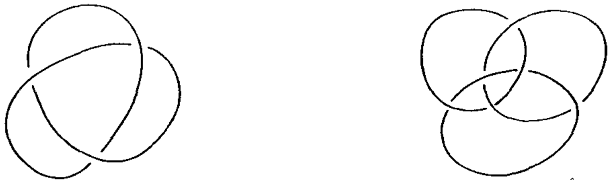
In this section we present a recent development in knot theory, which has turned out to be very closely related to classical combinatorial constructions like the Tutte polynomial of a graph. We will not have the opportunity to discuss another exciting development: the recent approach of Vassiliev (1990), which gives a systematic method to produce knot invariants, and reduces in a completely different way to combinatorial problems, see also Birman and Lin (1993). We refer to Burde and Zieschang (1985) for a broad development of the "classical" knot theory. More extensive surveys of the recent progress are Kauffman (1988), Lickorish (1988) and Birman (1993).

A *link* L with $c(L)$ *components* consists of $c(L)$ disjoint simple smooth closed curves in $\mathbb{R}^3$. A *knot* is a link with one component. The natural way to represent a link L is by means of a *link diagram*, which is obtained from L by projecting it onto a plane in such a way that the projection of each component is smooth and at most two curves intersect at any point. At each crossing point of the link diagram the curve which goes over the other is specified as shown in fig. 7.1.

Two links are *equivalent* if one can be deformed continously to the other in three dimensional space. A knot is *trivial* if it is equivalent to the knot without a crossing, the *unknot* "O". A link is trivial if it is equivalent to a disjoint union of trivial, unlinked knots, that is, to a link without a crossing. The *union* "$\cup$" operation used for this places two link diagrams into the plane so that they do not touch.

Clearly a link diagram represents a unique link up to equivalence, but many diagrams can represent the same link. Modifying a link diagram locally as represented in fig. 7.2 does not change the link represented; these local changes are known as the *Reidemeister moves* of types I, II and III. The classic theorem of combinatorial knot theory is:

Theorem 7.1. *Two link diagrams represent equivalent links if and only if one can be obtained from the other by a finite sequence of Reidemeister moves.*



The trefoil knot

The Borromean rings: a 3-component link

Figure 7.1.

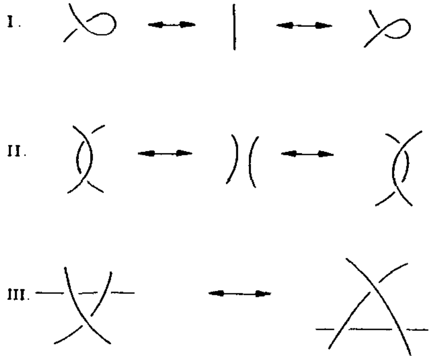


Figure 7.2.

Reidemeister's classic theorem is an existence theorem: it does not provide an algorithm of any "time-bounded" complexity for testing whether or not two links are equivalent, see for example the remark in Burde and Zieschang (1985).

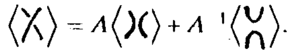
As a result, any invariant f of a link which is both easily calculated and has the property that L_1 and L_2 are equivalent only when $f(L_1) = f(L_2)$ is clearly of great significance in the theory of knots.

Here we shall present an easily derived "partial invariant", namely the Jones polynomial, discovered by Jones (1985). In order to do this we introduce first the closely related bracket polynomial introduced by Kauffman (1987).

The bracket polynomial

For any link diagram L define a Laurent polynomial $\langle L\rangle$ in one variable A, which obeys the following three rules:

(i) $\langle O\rangle = 1$,
(ii) $\langle L \cup O\rangle = -(A^2 + A^{-2})\langle L\rangle$,

(iii) 

Notes. (I) A *Laurent polynomial* in z is a polynomial featuring positive powers of z and z^{-1}.

(II) The rule (iii) is applied locally, that is, at each crossing of the link diagram.

(III) Rules (i) to (iii) recursively define the bracket $\langle L\rangle$ for every link L: (i) starts the recursion with the unknot, (iii) expresses $\langle L\rangle$ in terms of brackets of links with fewer crossings, and (ii) deletes components without crossings, for example trivial links.

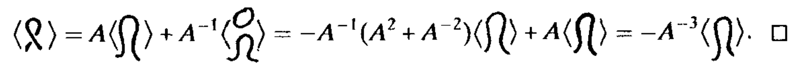
The fundamental properties of $\langle\cdot\rangle$ are summed up in:

Theorem 7.2. *For any link L the bracket polynomial $\langle L\rangle$ is independent of the order in which rules* (i) *to* (iii) *are applied to the crossings and furthermore is invariant under the Reidemeister moves* (II) *and* (III).

It is important to note:

Proposition 7.3. *The bracket polynomial is not invariant under Reidemeister move* (I).

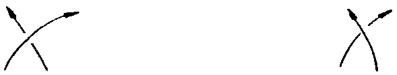
Proof. Apply move (I) to get



□

Oriented links and the Jones polynomial

Suppose now that the link L is *oriented*, by which we mean that each of its components is assigned an arrow representing a direction of motion along the given component. Once the orientation has been given we may assign a *sign* to each crossing of the link diagram by the rule displayed in fig. 7.3.



+ve crossing −ve crossing

Figure 7.3.

A crossing is *positive* if the over arc at the crossing is on the left as one approaches the crossing in the direction of the two indirected arcs.

The *writhe* of an oriented link L is the sum of the signs at the crossings of L and is denoted by $\omega(L)$ (see fig. 7.4). The writhe is *not* an invariant of a link since it changes by ±1 under the type I Reidemeister move. However, it turns out that if we combine the writhe with the bracket polynomial of Kauffman we get a link invariant.

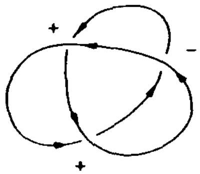


Figure 7.4. An oriented link with writhe $\omega(L) = 1$.

Theorem 7.4. *For an oriented link L, the function*

$$V_L(t) = f_L(t^{-1/4})$$

where

$$f_L(A) = (-A^3)^{\omega(L)}\langle L\rangle$$

is an invariant of the link.

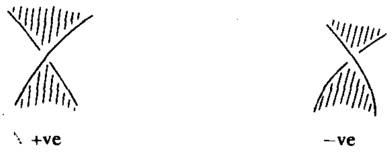
$V_L(t)$ is called the *Jones polynomial* of the oriented link L.

From knots to signed graphs

Now given any unoriented link L there is a natural way in which to associate with it a *signed graph* $G(L)$, constructed as follows.

Two-colour the faces of the link diagram of L with colours black and white in such a way that adjacent faces have different colours. (This is possible since L, regarded as a graph, is planar and regular of degree 4, hence Eulerian. Thus its dual graph is bipartite.) By convention let the unbounded face be white. Let $G(L)$ have vertices corresponding to the black faces and join two vertices of $G(L)$ when the corresponding faces are the opposite faces of a crossing. The sign of the crossing is defined by the convention illustrated in fig. 7.5. This is that a crossing is *positive* when viewed along the edge joining the two black faces the edge passing over at the crossing is on the left; otherwise it is *negative*.

In general the graph $G(L)$ will have both positive and negative edges but there is an important class of knots (links), called *alternating links*, for which $G(L)$ has edges of only one sign.


+ve

−ve

Figure 7.5.

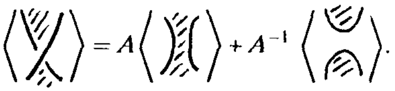
A link diagram L is *alternating* if the crossings alternate under–over–under–over... as the link is traversed. It is very easy to verify. A link is *alternating* if it has some link diagram that is alternating.

Lemma 7.5. *A link diagram L is alternating iff the edges of the associated signed graph $G(L)$ are all positive or all negative.*

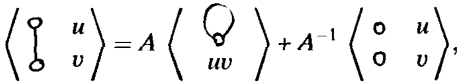
Alternating links; the Jones, bracket and Tutte polynomial

In general the bracket polynomial and Jones polynomial of a link will not be completely specified by the graph $G(L)$, but will also depend on the signs associated with the edges of $G(L)$. However, in the important case where L is an alternating link diagram, we saw above that all edges of $G(L)$ have the same sign and hence essentially the undirected graph $G(L)$ determines L.

Now consider the fundamental bracket equation or skein diagram (as it is called in knot theory),



In terms of the associated graph $G(L)$, this can be translated to



which is exactly the contract/delete formulation used in the definition of the Tutte polynomial of a graph and matroid (see chapter 9).

It is not surprising therefore that we have the following fundamental relationship.

Theorem 7.6. *If L is an oriented alternating link diagram and G denotes its associated unsigned "black-face" graph then the Jones polynomial $V_L(t)$ is given by*

$$V_L(t) = (t^{-1/4})^{3\omega(L)-2}T(G;-t,-t^{-1})$$

where $T(G;x,y)$ is the Tutte polynomial of G.

In other words, the Jones polynomial of an alternating link diagram is given by the evaluation of the Tutte polynomial of its associated black-face graph along the hyperbola $xy = 1$, up to an easily derived factor.

A proof of a conjecture by Tait

As an example of the use of these knot polynomials we show a very simple proof of a longstanding conjecture by Tait (1898).

Consider first the knot K of fig. 7.6. The crossing joining AB is called an *isthmus* because in the associated link diagram the edge AB will be an isthmus in the usual graph-theoretic sense. A link diagram is *reduced* if it has no crossing which is an isthmus.

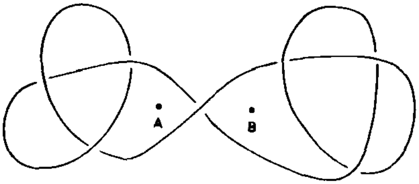


Figure 7.6.

It is easy to see that the knot represented by fig. 7.6 is equivalent to the alternating knot K' obtained by "twisting one component and removing the isthmus crossing AB".

The key fact we shall need is the following, obtained by considering the representation of $\langle K\rangle$ in terms of the Tutte polynomial of $G(K)$:

Proposition 7.7. *Let L be a reduced alternating diagram, then in the bracket polynomial $\langle K\rangle$*

$$\max \text{degree}\langle K\rangle = V + 2(W-1),$$

$$\min \text{degree}\langle K\rangle = -V - 2(B-1),$$

where V is the number of regions and W and B are the numbers of white and black regions respectively in the shaded graph $G(L)$.

We now state and prove Tait's conjecture.

Theorem 7.8 (Murasugi 1987, Thistlethwaite 1987). *The number of crossings in a reduced alternating projection of a link L is a topological invariant of L.*

Expressed more informally, this means that if we have a link diagram which is alternating and contains no isthmuses and has n crossings then we know that there is no other reduced alternating link diagram representing the same knot and having a different number of crossings.

Proof. Let span(L) denote the difference between the maximum and minimum degrees of A in the bracket polynomial $\langle L\rangle$. By Theorem 7.7 (with the notation introduced there) we have

$$\max \text{degree}\langle K\rangle = V + 2(W-1),$$

$$\min \text{degree}\langle K\rangle = -V - 2(B-1).$$

Thus span$(L) = 2V + 2(W + B - 2)$ and since $W + B = V + 2$ (Euler's formula for planar graphs) we have span$(L) = 4V$. But $f(L) = a^{-\omega(L)}\langle L\rangle$ is a topological invariant of links, thus span(L) is also, and hence V, the number of crossings in an alternating presentation, is also a topological invariant. □

References

Alon, N., and V.D. Milman

[1983] Embedding of l^k_∞ in finite dimensional Banach spaces, *Israel J. Math.* **45**, 265–280.

Arvola, W.A.

[1992] The fundamental group of the complement of an arrangement of complex hyperplanes, *Topology* **31**, 757–765.

Avis, D., and M. Deza

[1991] L^1-embeddability, complexity and multicommodity flows, *Networks* **21**, 565–617.

Babai, L.

[1981] On the order of uniprimitive permutation groups, *Ann. of Math.* **113**, 553–568.

[1982] On the order of doubly transitive permutation groups, *Invent. Math.* **65**, 473–484.

Baclawski, K.

[1981] Rings with lexicographic straightening law, *Adv. in Math.* **39**, 185–213.

Baclawski, K., and A.M. Garsia

[1981] Combinatorial decompositions of rings, *Adv. in Math.* 39, 155–184.

Barahona, F., and A.R. Mahjoub

[1986] On the cut polytope, *Math. Programming* **36**, 157–173.

Birman, J.S.

[1993] New points of view in knot theory theory, *Bull. Amer. Math. Soc.* **28**, 253–287.

Birman, J.S., and X.-S. Lin

[1993] Knot polynomials and Vassiliev's invariants, *Invent. Math.* **111**, 225–270.

Björner, A.

[1982] On the homology of geometric lattices, *Algebra Universalis* **14**, 107–128.

Björner, A., and M. Wachs

[1982] Bruhat order of Coxeter groups and shellability, *Adv. in Math.* **43**, 87–100.

Björner, A., and G.M. Ziegler

[1991] Broken circuit complexes: factorizations and generalizations, *J. Combin. Theory B* **51**, 96–126.

[1992] Combinatorial stratification of complex arrangements, *J. Amer. Math. Soc.* **5**, 105–149.

Björner, A., A.M. Garsia and R.P. Stanley

[1982] An introduction to Cohen–Macaulay partially ordered sets, in: *Ordered Sets*, ed. I. Rival (Reidel, Dordrecht) pp. 583–615.

Björner, A., M. Las Vergnas, B. Sturmfels, N.L. White and G.M. Ziegler

[1993] *Oriented Matroids, Encyclopedia Math.*, Vol. 46 (Cambridge University Press, Cambridge).

Bochert, A.

[1889] Über die Zahl der verschiedenen Werthe, die eine Funktion gegebener Buchstaben durch Vertauschung derselben erlangen kann, *Math. Ann.* **33**, 584–590.

Burde, G., and H. Zieschang

[1985] *Knots* (Walter de Gruyter, Berlin).

Cordovil, R.

[1994] On the homotopy of the Salvetti complexes determined by simplicial arrangements, *European J. Combin.* **15**, 207–215.

De Concini, C., and V. Lakshmibai

[1981] Arithmetic Cohen–Macaulayness and arithmetic normality for Schubert varieties, *Amer. J. Math.* **103**, 835–850.

De Concini, C., and C. Procesi

[1981] Hodge algebras, a survey, in: *Proc. Int. Conf. on Young Tableaux and Schur Functors in Algebra and Geometry, Torun, Poland, 1980, Astérisque* **87–88**, 79–83.

De Concini, C., D. Eisenbud and C. Procesi

[1982] Hodge algebras, *Astérique* **91**.

Deligne, P.

[1972] Les immeubles des groupes de tresses généralisés, *Invent. Math.* **17**, 273–302.

Deza, M., and M. Laurent

[1992a] Facets for the complete cut cone I, *Math. Programming* **56**, 121–160.

[1992b] Facets for the complete cut cone II: clique-web inequalities, *Math. Programming* **56**, 161–188.

[1992c] New results on facets of cut cone, *J. Combin. Inform. System Sci.* 17, 19–38.

Deza, M., VP. Grishukhin and M. Laurent

[1993] The hypermetric cone is polyhedral, *Combinatorica* **13**, 397–411.

Deza, M. (M.E. Tylkin)

[1960] On the Hamming geometry of unit cubes, *Dokl. Akad. Nauk SSSR* **134**, 1037–1040.

Dvoretzky, A.

[1959] A theorem on convex bodies and applications to Banach spaces, *Proc. Nat. Acad. Sci. USA* **45**, 223–226.

Edelman, P.H., and V. Reiner

[1993] A counterexample to Orlik's conjecture, *Proc. Amer. Math. Soc.* **118**, 927–929.

[1994] Free hyperplane arrangements between A_{n-1} and B_n, *Math. Z.* **215**, 347–365.

Eisenbud, D.

[1980] Introduction to algebras with straightening laws, in: *Ring Theory and Algebra, Proc. 3rd Oklahoma Conf.*, ed. B.R. McDonald (Marcel Dekker, New York).

Falk, M.J.

[1993] Homotopy types of line arrangements, *Invent. Math.* **111**, 139–150.

Falk, M.J., and R. Randell

[1985] The lower central series of fiber-type arrangement, *Invent. Math.* **82**, 77–88.

Figiel, T., J. Lindenstrauss and V. Milman

[1977] The dimension of almost spherical sections of convex bodies, *Acta Math.* **139**, 53–94.

Garsia, A.M.

[1980] Combinatorial methods in the theory of Cohen–Macaulay rings, *Adv. in Math.* **38**, 229–266.

Grötschel, M., L. Lovász and A. Schrijver

[1988] *Geometric Algorithms and Combinatorial Optimization, Algorithms and Combinatorics*, Vol. 2 (Springer, Berlin).

Halmos, P.

[1950] *Measure Theory* (Van Nostrand, Princeton, NJ).

Hattori, A.

[1975] Topology of $\mathbb{C}^n$ minus a finite number of hyperplanes in general position, *J. Fac. Sci. Univ. Tokio* **22**, 205–219.

Hochster, M.

[1972] Rings of invariants of tori, Cohen–Macaulay rings generated by monomials, and polytopes, *Ann. of Math. (2)* **96**, 318–337.

Iri, M.

[1970] On an extension of the max-flow–min-cut theorem to multicommodity flows, *J. Oper. Res. Japan* **13**, 129–135.

Jambu, M., and H. Terao

[1984] Free arrangements of hyperplanes and supersolvable lattices, *Adv. in Math.* **52**, 248–258.

[1989] Arrangements of hyperplanes and broken-circuits, in: *Singularities, Proc. Int. Conf. on Singularities, Iowa City, 1986*, ed. R. Randell, *Contemporary Math.* **90**, 147–162.

Johnson, V.B., and G. Schechtman

[1981] On subspaces of L_1 with maximal distance to euclidean space, in: *Proc. Res. Workshop on Banach Space Theory*, ed. Bor Luh Lin (University of Iowa) pp. 83–96.

Jones, V.F.R.

[1985] A polynomial invariant for knots via von Neumann algebras, *Bull. Amer. Math. Soc.* **12**, 103–111.

Jordan, C.

[1961] *Oeuvres I*, ed. J. Dieudonné (Gauthier-Villars, Paris) pp. 408, 453.

Karzanov, A.V.

[1985] Metrics and undirected cuts, *Math. Programming* **32**, 183–198.

Kauffman, L.H.
[1987] State models and the Jones polynomial, *Topology* **26**, 395–407.
[1988] New invariants in the theory of knots, *Amer. Math. Monthly* 95, 195–242.

Kelly, J.B.
[1970] Metric inequalities and symmetric differences, in: *Inequalities II* (Academic Press, New York) pp. 193–212.
[1975] Hypermetric spaces, in: *Lecture Notes in Mathematics*, Vol. 490 (Springer, Berlin) pp. 17–31.

Kind, B., and P. Kleinschmidt
[1979] Schälbare Cohen–Macaulay-Komplexe und ihre Parametrisierung, *Math. Z.* **167**, 173–179.

Kunz, E.
[1985] *Introduction to Commutative Algebra and Algebraic Geometry* (Birkhäuser, Boston).

Lickorish, W.B.R.
[1988] Polynomials for links, *Bull. London Math. Soc.* **20**, 558–588.

Liebeck, M.V.
[1984] On minimal degrees and base sizes of primitive permutation groups, *Arch. Math.* **43**, 11–15.

Liebeck, M.W., and J. Saxl
[1991] Minimal degrees of primitive permutation groups with an application to monodromy groups of covers of Riemann surfaces, *Proc. London Math. Soc.* **63**, 266–314.

Lovász, L., and M.D. Plummer
[1986] *Matching Theory, Ann. Discrete Math.* **29**.

Milman, V.
[1982] Some remarks about embedding of l_1^k in finite dimensional spaces, *Israel J. Math.* **43**, 129–138.

Murasugi, K.
[1987] Jones polynomials and classical conjectures in knot theory, *Topology* **26**, 187–194.

Onaga, K., and O. Kakusho
[1971] On feasibility conditions of multicommodity flows in networks, *IEEE Trans. Circuit Theory* **CT-18**, 425–429.

Orlik, P.
[1989] *Introduction to Arrangements, CBMS Regional Conf. Ser. Math.* **72**.

Orlik, P., and L. Solomon
[1980] Combinatorics and topology of complements of hyperplanes, *Invent. Math.* **56**, 167–189.

Orlik, P., and H. Terao
[1992] *Arrangements of Hyperplanes, Grundlehren der Mathematischen Wissenschaften*, Vol. 300 (Springer, Berlin).

Paris, L.
[1993a] Universal cover of Salvetti's complex and topology of simplicial arrangements of hyperplanes, *Trans. Amer. Math. Soc.* **340**, 149–178.
[1993b] The Deligne complex of a real arrangement of hyperplanes, *Nagoya Math. J.* **131**, 39–65.

Pyber, L.
[1993] On the orders of doubly transitive permutation groups, elementary estimates, *J. Combin. Theory A* **62**(2), 361–366.

Rose, L.L., and H. Terao
[1990] Hilbert polynomials and geometric lattices, *Adv. in Math.* **84**, 209–225.

Rota, G.-C., and L.H. Harper
[1971] Matching theory, an introduction, in: *Advances in Probability Theory and Related Topics I*, ed. P. Ney (Marcel Dekker, New York) pp. 169–215.

Saito, K.
[1980] Theory of logarithmic differential forms and logarithmic vector fields, *J. Fac. Sci. Univ. Tokyo, Sci. IA* **27**, 265–291.

Salvetti, M.
[1987] Topology of the complement of real hyperplanes in $\mathbb{C}^N$, *Invent. Math.* **88**, 603–618.
[1993] On the homotopy theory of the complexes associated to metrical-hemisphere complexes, *Discrete Math.* **113**, 155–177.

Shafarevich, I.R.

[1977] *Basic Algebraic Geometry* (Springer, New York).

Solomon, L., and H. Terao

[1987] A formula for the characteristic polynomial of an arrangement, *Adv. in Math.* **64**, 305–325.

Spanier, E.H.

[1966] *Algebraic Topology* (McGraw-Hill/Springer).

Stanley, R.P.

[1972] Supersolvable lattices, *Algebra Universalis* **2**, 197–217.

[1975] Cohen–Macaulay rings and constructible polytopes, *Bull. Amer. Math. Soc.* **81**, 133–135.

[1984] T-free arrangements of hyperplanes, in: *Progress in Graph Theory*, eds. J.A. Bondy and U.S.R. Murty (Academic Press, New York) p. 539.

[1986] *Enumerative Combinatorics*, Vol. I (Wadsworth and Brooks/Cole, Monterey, CA).

Sturmfels, B., and N. White

[1989] Gröbner bases and invariant theory, *Adv. in Math.* **76**, 245–259.

[1991] Computing combinatorial decompositions of rings, *Combinatorica* **11**, 275–293.

Tait, P.G.

[1898] On knots I, II, III, in: *Scientific Papers*, Vol. I (Cambridge University Press, London) pp. 273–347.

Terao, H.

[1980] Arrangements of hyperplanes and their freeness I, *J. Fac. Sci. Univ. Tokyo Sci. IA* **27**(2) 293–312.

[1981] Generalized exponents of a free arrangement of hyperplanes and Shephard–Todd–Brieskorn formula, *Invent. Math.* **63**, 159–179.

[1986] Modular elements of lattices and topological fibration, *Adv. in Math.* **62**, 135–154.

Thistlethwaite, M.B.

[1987] A spanning tree expansion of the Jones polynomial, *Topology* **26**, 297–209.

Vassiliev, V.A.

[1990] Cohomology of knot spaces, in: *Theory of Singularities and its Applications*, ed. V.I. Arnol'd, *Advances in Soviet Mathematics*, Vol. 1 (AMS, Providence, RI) pp. 23–69.

Wielandt, H.

[1964] *Finite Permutation Groups* (Academic Press, New York).

Yuzvinsky, S.

[1993] The first two obstructions to the freeness of arrangements, *Trans. Amer. Math. Soc.* **335**, 231–244.

Zemlyachenko, V.N., N.M. Vornienko and R.I. Tyshkevich

[1985] Graph isomorphism problems, *J. Soviet Math.* **29**, 1426–1481.

Ziegler, G.M.

[1989] Combinatorial construction of logarithmic differential forms, *Adv. in Math.* **76**, 116–154.

[1990] Matroid representations and free arrangements, *Trans. Amer. Math. Soc.* **320**, 525–541.

[1993] On the difference between real and complex arrangements, *Math. Z.* **212**, 1–11.

Ziegler, G.M., and R.T. Živaljević

[1993] Homotopy types of arrangements via diagrams of spaces, *Math. Ann.* **295**, 527–574.

Part V
Horizons

1. Introduction

If pressed hard to give a rigorous classification of the subject matter, I would say that infinite combinatorics is part of set theory and hence a part of mathematical logic.

On the other hand, the problems and ideas of this topic are clearly of combinatorial character. Quite a few of the basic questions arose by simply replacing integer parameters by infinite cardinals in finite problems. It is however a fact that most of the good problems lead to situations which can only be handled adequately by the methods of logic developed in the last three decades: forcing, large cardinals, constructibility and its generalizations.

As to the selection of the material presented in this chapter there is clearly no way to achieve completeness. Infinite combinatorics could fill another book or two. I will try to illustrate the phenomena described above on a few carefully selected problems, which I hope will all sound familiar and natural to experts of finite combinatorics. I will prove some results of combinatorial character on these problems and then state the independence results required to fill the gaps in our knowledge.

I would like to mention that a chapter describing infinite combinatorics written more in the spirit of set theory appeared in the Handbook of Mathematical Logic (see Kunen 1977).

I am not going to present a specific axiom system of set theory, but it should be emphasized that anything we will do can be done in the well-known Zermelo–Fraenkel axiom system of set theory with *the axiom of choice*. If we use any other set-theoretical assumptions, these will be stated explicitly. I will list some notation, elementary facts and theorems without proofs in the Appendix. These, I hope, will be mostly self-explanatory and (or) well known to most of the readers. Still I will give reference to these results in case I use them in sections 2–6. For a detailed development of set theory from the axiom systems, see, e.g., Jech (1978).

2. Infinite graphs and hypergraphs. Analysis of a very simple problem

Quite a few definitions and results given for finite graphs, digraphs and hypergraphs, extend to infinite ones, without essential modification. This is precisely what we are going to do. If we do not define a concept in the infinite case, or if we use a result, valid in the finite case, without any further argument this means that the original definition or argument applies without any change. We will only consider simple graphs, which we will simply call graphs.

The first problems arise if we consider some invariant number associated with an object. Clearly, in general, we intend to replace the number of elements of a set by the cardinality of the set. Infinite cardinals are well-ordered by order of magnitude, hence no problem arises if we must take the *minimum* of certain numbers. So, for example, $\chi(G)$, $\bar{\chi}(G)$, $\gamma(G)$, and $\delta(G)$ are defined without any change.

An infinite set of cardinals does not necessarily have a largest member, so for example the definition of $\alpha(G), \omega(G)$, and $\Delta(G)$ does not extend automatically, and there are indeed at least two different ways to define them. For example, we could define, for $G = (V, E)$:

$$\alpha(G) = \sup|\{|X| : X \subset V, X \text{ is a stable set in } G\}\ .$$

This has two advantages. First, for finite graphs it coincides with the old definition, and second, it is what everybody would do, to start with. However, there is a definite disadvantage, as well. If, say, $\alpha(G) = \omega$, we do not know if there is an infinite stable set, or all stable sets are finite, but unbounded in size. We could define $\hat{\alpha}(G)$ as the *smallest* cardinal κ such that there is no stable set of size κ. This can also be expressed as a supremum, namely

$$\hat{\alpha}(G) = \sup\{|X|^+ : X \subset V, X \text{ is a stable set in } G\}\ .$$

Now $\hat{\alpha}(G) = \alpha(G) + 1$ in case G [or even $\alpha(G)$] is finite. $\hat{\alpha}(G) = \omega$ if there are only finite stable sets, but the size of them is unbounded, and $\hat{\alpha}(G) = \omega_1$ in case the largest independent set is of size $\aleph_0$. So the second possibility is technically much better. Not to hurt the feelings of those (too much) who are willing to consider countable stable sets, but refuse to think of ω_1, we will use both notations for α, and for any other invariant defined similarly. For example, $\hat{\Delta}(G) = \omega$ means that the degree of every vertex is finite, but there is no finite upper bound. To see this technique work, let us try to find the infinite analogue of a very simple theorem: *If G is finite, then $\chi(G) \leq \Delta(G) + 1$* (cf. chapter 4).

Theorem 2.1. *$\chi(G) \leq \hat{\Delta}(G)$ holds for every graph G.*

Proof. Set $\hat{\Delta}(G) = \kappa$. Let $V = \{x_\beta : \beta < \alpha\}$ be a one-to-one enumeration of V_j; we will call it a *well-ordering* of V. We want to define stable sets $(V_\xi : \xi < \kappa)$ such that $\bigcup_{\xi<\kappa} V_\xi = V$. We apply transfinite recursion. Assume $\beta < \alpha$, and that for $\gamma < \beta$ we already decided which V_ξ contains x_γ, i.e., there is a function ξ defined for all $\gamma < \beta$ with the correspondence $\xi(\gamma) = \xi \leftrightarrow x_\gamma \in V_\xi$. Now the degree of x_β is less than κ, hence there is a $\xi < \kappa$ such that G has no edge $\{x_\gamma, x_\beta\}$ with $\xi(\gamma) = \xi$. Set $\xi(\beta) = \xi$ for the minimal ξ of this kind. By Theorem A.11 (ii), this defines $\xi(\beta)$ for every $\beta < \alpha$, and the sets V_ξ for $\xi < \kappa$. Clearly the sets V_ξ are stable and cover V. □

The question arises if Theorem 2.1 is best possible. We know this for finite κ, as is shown by K_κ. Assume κ is an infinite limit cardinal. Then, for a G which is the disjoint union of K_λ's for $\lambda < \kappa$, $\hat{\Delta}(G) = \chi(G) = \kappa$, hence the result is best possible in this case too.

Assume now $\hat{\Delta}(G)$ is an infinite successor cardinal $\kappa = \lambda^+$. $\hat{\Delta}(K_{\lambda^+}) = \lambda^{++}$ and we have no immediate example. In fact, in this case the result can be improved. First we prove:

Lemma 2.2. *Assume G is a connected graph and $\Delta(G) \leq \kappa$ for some $\kappa \geq \omega$. Then $|V| \leq \kappa$.*

Proof. Let $x \in V$ be arbitrary, and, for $i < \omega$, let

$$V_i = \{y \in V\colon \text{the distance between } x \text{ and } y \text{ in } G \text{ is } i\} .$$

$V = \bigcup_{i<\omega} V_i$, since G is connected. Now we prove by induction on $i < \omega$, that $|V_i| \leq \kappa$. This is true for $i = 0$, since $V_0 = \{x\}$, $|V_0| = 1 \leq \kappa$. Assume $|V_i| \leq \kappa$ for some i. Since each element of V_{i+1} is adjacent to some element of V_i, and $\Delta(G) \leq \kappa$, it follows that $|V_{i+1}| \leq |V_i| \cdot \kappa \leq \kappa^2 = \kappa$. Hence $|V_i| \leq \kappa$ holds for $i < \omega$. But then

$$|V| \leq \kappa \cdot \omega = \kappa . \qquad \square$$

Corollary 2.3. *Assume $\Delta(G) \leq \kappa$ for some $\kappa \geq \omega$. Then $\chi(G) \leq \kappa$. As a consequence of this, if $\hat{\Delta}(G) = \kappa^+ > \omega$ for some G, then*

$$\chi(G) \leq \kappa < \hat{\Delta}(G) .$$

Proof. By Lemma 2.2, each component of G has cardinality at most κ. Clearly, V is the union of κ sets, each having at most one element in common with each component. But all sets of this kind are obviously stable. $\square$

Now one could argue that the proof of Theorem 2.1 we gave is superfluous. Indeed, if $\hat{\Delta}(G) = \kappa > \omega$, then Corollary 2.3 obviously implies Theorem 2.1. The case $\hat{\Delta}(G) = \omega$ easily reduces to the cases $\hat{\Delta}(G) < \omega$. In case $\hat{\Delta}(G) = \kappa < \omega$, we know the result if G is finite. The question arises if this fact alone implies the result for every G. The answer is affirmative, as stated first in this form by de Bruijn and Erdős (1951), but the proof of this is certainly more involved than the proof of Theorem 2.1. The main aim of the next section is to isolate and prove the general principles, called *compactness* arguments, yielding this type of results.

3. König's lemma and compactness

The theorem of König (1927) is probably the earliest result and certainly the earliest well-known result, about infinite graphs. To state it, we have to define the concept of an infinite path.

A graph $G = (V, E)$ is a *one-way infinite path* if there is a well-ordering $\{x_i\colon i < \omega\} = V$ of type ω of the vertex set, such .hat $E = \{\{x_i, x_{i+1}\}\colon i < \omega\}$. The isomorphism type of a one-way infinite path is denoted by P_∞.

A graph $G = (V, E)$ is a *two-way infinite path* if $V = \{x_p\colon p \in \mathbb{Z}\}$ is a one-to-one enumeration of the vertex set, and $E = \{\{x_p, x_{p+1}\} : p \in \mathbb{Z}\}$. The isomorphism type of a two-way infinite path is denoted by $P_{\omega,\omega}$. We now prove König's lemma.

Theorem 3.1. *Assume G is an infinite connected graph with $\hat{\Delta}(G) \leqslant \omega$. Then $P_\omega \subseteq G$, i.e., G contains a one-way infinite path.*

Proof. We are going to define a sequence $(x_i : i < \omega)$ of vertices, and a sequence $(V_i : i < \omega)$ of vertex sets, by recursion on $i < \omega$. Let $V_0 = V$, and let x_0 be an arbitrary vertex of G. Assume $V_0 \supset \cdots \supset V_i$ and $x_0 \in V_0, \ldots, x_i \in V_i$ are already defined in such a way that $G[V_i]$ is connected and infinite, and $x_j \not\in V_i$ for $j < i$. By the assumption $\hat{\Delta}(G) \leqslant \omega$, $N(x_i)$ is finite. Since $G[V_i]$ is connected, each component of $G[V_i] - x_i$ contains an element of $N(x_i)$. It follows that $G[V_i] - x_i$ has only finitely many components, hence one of them is infinite. Call one infinite component V_{i+1}, and let x_{i+1} be an element of $N(x_i) \cap V_{i+1}$. Then $V_{i+1} \subset V_i$, $G[V_{i+1}]$ is connected and infinite, and $x_j \not\in V_{i+1}$ for $j < i + 1$. Hence x_i and V_i are defined for all $i < \omega$. By the construction, $x_j \neq x_i$ for $j < i < \omega$ and x_i is adjacent to x_{i+1} in G. □

Traditionally, König's lemma is stated for trees. This does not make any difference, since, by Zorn's lemma, a connected graph has a spanning tree.

A pair $(T, <)$ is a said to be a *set theoretic tree*, briefly an *S-tree*, if $<$ is a partial order on T, i.e., irreflexive and transitive on T, and for all $x \in T$, the set $\hat{x} = \{y \in T : y < t)$ is well-ordered by $<$. For $x \in T$, let $\alpha_T(x) = \alpha(x) = \operatorname{typ} \hat{x}(<)$; $T_\alpha = \{x \in T : \alpha(x) = \alpha\}$ is the αth level of T, $h(T) = \min\{\alpha : T_\alpha = \emptyset\}$. $h(T)$ is called the *height* of T. Clearly $T = \bigcup_{\alpha < h(T)} T_\alpha$. $B \subset T$ is a *branch* of T, if B is ordered by $<$ and $B \cap T_\alpha \neq \emptyset$ for all $\alpha < h(T)$.

Definitions 3.2. Let $\kappa \geqslant \omega$. An S-tree $(T, <)$ is said to be a *κ S-tree* if $h(T) = \kappa$, and $|T_\alpha| < \kappa$ for $\alpha < h(T)$.

A cardinal κ has the *tree property* if every κ S-tree has a branch.

It is a much investigated problem to determine which cardinals have the tree property. For example, it can be proven that ω_1 does not have the tree property. The examples showing this are called Aronszajn trees (see, e.g., Kunen 1977, p. 384). We only gave here this definition to reformulate König's lemma. However, our study of the generalizations of Ramsey's theorem will lead us back to this property.

Theorem 3.3. *ω has the tree property.*

Proof. Let $(T, <)$ be an ω S-tree. We may assume it has a root, i.e., $|T_0| = 1$. We define a graph $G = (T, E)$. E consists of pairs $\{x, y\}$ of the form $x \in T_i$, $y \in T_{i+1}$, $x < y$, for $i < \omega$. Since $|T_i| < \omega$ for $i < \omega$, we clearly have $\hat{\Delta}(G) \leqslant \omega$. Since, by the S-tree property, for all $y \in T_{i+1}$ there is exactly one $x \in T_i$ with $x < y$, the existence of an infinite path in G clearly implies the existence of a branch of T. □

To see that Theorem 3.3 is really a restatement of König's lemma, we deduce Theorem 3.1 from Theorem 3.3. By the remark made after the proof of Theorem 3.1 we may assume that we are given a graph $G = (V, E)$ with $\hat{\Delta}(G) \leqslant \omega$ which is a tree. Let x be an arbitrary element of V. For each $y, z \in V$, set $z < y$ if and only if $z \neq y$ and z lies on the unique path connecting x and y. It is easy to verify that $(V, <)$ is an ω S-tree, and every branch of $(V, <)$ is an infinite path of G.

We are after a generalization of König's lemma, which would yield the desired compactness results. We already know that the κ-tree property is too strong to be true for every κ. We give another reformulation of Theorem 3.1.

Theorem 3.4. *For each $n < \omega$, let A_n be a non-empty finite set. Let $A = \times_{n \in \omega} A_n$ and $B_n = \times_{i<n} A_i$ for $n \in \omega$. Assume we are given a sequence $f_n \in B_n$ for $n < \omega$. Then there exists an $f \in A$, such that for all $i \in \omega$ there is an $n, i \leqslant n < \omega$, with $f \restriction i = f_n \restriction i$.*

Proof. Let $T_i = \{g \in B_i\colon$ there is an $n \geqslant i$ with $f_n \restriction i = g\}$. Clearly, $T_i \subset B_i$ for $i < \omega$. Let $T = \bigcup_{i<\omega} T_i$. Write $g < h \Leftrightarrow g \subset h$ for $g, h \in T$. Since the elements of T are functions, $(T, <)$ is an S-tree. Since every section of a $g \in T$ belongs to T, T_i is the ith level of T, and $T_i \subset B_i$ implies that T_i is finite; hence T is an ω S-tree. By Theorem 3.3 T has a branch B. Thus, $\bigcup B = f$ is a function, $f \in A$ and f satisfies the requirements of Theorem 3.4. □

I leave it to the reader to see that Theorem 3.4 just as easily implies Theorem 3.3. Hence, this is just another restatement of König's lemma. However, this one can easily be generalized to the right theorem.

Let $[X]^{<\lambda}$ (respectively $[X]^{\lambda}$) denote the subsets of X of size less than λ (size λ). The following is called *Rado's selection lemma*.

Theorem 3.5. *Let α be an arbitrary index set, let A_β be a finite non-empty set for every $\beta < \alpha$, and let $A = \times_{\beta<\alpha} A_\beta$ be the Cartesian product of the A_β's. For each $V \in [\alpha]^{<\omega}$, set $B_V = \times_{\beta \in V} A_\beta$, and let $f_V \in B_V$ be given. Then there exists a choice function $f \in A$ such that for all $W \in [\alpha]^{<\omega}$ there is a $W \subseteq V \in [\alpha]^{<\omega}$ such that*

$$f \restriction W = f_V \restriction W .$$

Proof. For $W \in [\alpha]^{<\omega}$, let $T_W = \{g \in B_W\colon$ there is a $V \in [\alpha]^{<\omega}$, $W \subseteq V$ such that $g = f_V \restriction W\}$. Note that $T_W \subseteq B_W$; hence T_W is finite. Let $A_W = \{f \in A : f \restriction W \in T_W\}$.

We have to prove that there is an f in the intersection of all A_W's. Choose A as the underlying set, and notice that the system

$$\mathcal{F} = \{A_W : W \in [\alpha]^{<\omega}\}$$

has the finite intersection property. Indeed, let $W_0, \ldots, W_{n-1}$ be finite subsets of α. Then $V = \bigcup_{i<n} W_i$ is finite. Since the sets A_β are non-empty, f_V has an extension f in A, $f_V \restriction W_i \in T_{W_i}$ for $i < n$, hence $f \in A_{W_i}$ for $i < n$.

By Theorem A.14, $\mathcal{F}$ can be extended to an ultrafilter U in A. For each $\beta < \alpha$ and $x \in A_\beta$, let $A_{\beta,x} = \{f \in A : f(\beta) = x\}$. Clearly, $A = \bigcup_{x \in A_\beta} A_{\beta,x}$ for each $\beta < \alpha$, and the summands are pairwise disjoint. U being an ultrafilter, and A_β being finite for each $\beta < \alpha$, there is a unique $f(\beta) \in A_\beta$ such that

$$A_{\beta,f(\beta)} \in U .$$

We claim that this $f \in A$ satisfies the requirements of the theorem. Let $W \in [\alpha]^{<\omega}$. Then, $\bigcap_{\beta \in W} A_{\beta,f(\beta)} \in U$, by the finite intersection property of U, and $\bigcap_{\beta \in W} A_{\beta,f(\beta)} \cap A_W \in U$ because of $A_W \in \mathcal{F}$. This implies that $f \restriction W \in T_W$, hence $f \in A_W$. □

Statements implying Rado's selection lemma were formulated earlier in other branches of mathematics, such as *Tychonov's product theorem* stating that the topological product of compact topological spaces is compact, and *Gödel's compactness theorem* stating that a set of formulas of a first order language has a model, provided every finite subset of these formulas has a model (see, e.g., Kunen 1977, p. 10). It is especially easy to deduce the selection lemma from Tychonov's theorem, since the finite sets A_β are compact in the discrete topology, and the sets A_W are closed in their product space. It is just as easy for logicians to see the derivation from Gödel's theorem, and also this is the case with the applications of the selection lemma. That is why consequences of the selection lemma are stated in the literature with the laconic remark "by compactness", and we will follow this practice in the future, but now we really deduce the de Bruijn–Erdős theorem, already quoted in section 2, from the selection lemma.

Theorem 3.6. *If $\chi(G') \leqslant k < \omega$ for every finite subgraph G' of G, then $\chi(G) \leqslant k$.*

Proof. Let $G = (\alpha, E)$, and let $A_\beta = k$ $(=\{0, \ldots, k-1\})$ for every $\beta < \alpha$. Using the notation of the selection lemma, let $f_V \in B_V$ be a good coloring of $G[V]$ for $V \in [\alpha]^{<\omega}$. Let f be the function given by the selection lemma. Clearly, f is a k-coloring of G. f is a good coloring of G, since for every edge $e = \{\beta, \gamma\}$ there is a $V \supseteq e$, such that $f \restriction e = f_V | e$, and hence $f(\gamma) = f_V(\gamma) \neq f_V(\beta) = f(\beta)$. □

The question arises whether or not the selection lemma can be true for
cardinals
with the word finite replaced by $<\kappa$ everywhere. This
large cardinals. Such a cardinal κ is called *strongly*
is immensely large. We will briefly speak about large
l cardinals are not strongly compact, but the problem
ces (or of proving some weak forms) remains. This is
ıly illustrate it by stating a few results concerning the
roved in Erdős and Hajnal (1968) that there exist
ith $\chi(G) = \kappa^+$, such that all subgraphs of cardinality
umber at most κ, for $\kappa \geqslant \omega$. We will prove this in
in the constructible universe, for every regular κ,

there is a graph of size κ which is $\aleph_1$ chromatic and all its subgraphs of size $<\kappa$ are $\leq\aleph_0$ chromatic if and only if κ does not have the tree property.

Finally, we mention the problem, due to Erdős and Hajnal, of the chromatic number jumping two or more cardinals. We state a special case.

Question 3.7. Assume the continuum hypothesis (see the Appendix). Does there exist a graph G with $\aleph_2$ vertices and $\chi(G) = \aleph_2$ such that all subgraphs of size $<\aleph_2$ have chromatic number at most $\aleph_0$?

It was proved in Baumgartner (1984) that an affirmative answer to this problem is consistent, and Foreman and Laver (1988) proved that it is consistent, relative to the consistency of a very large cardinal, that the answer is negative. If in Problem 3.7 $\aleph_2$ can be replaced by $\aleph_3$, or by anything greater than $\aleph_2$ and smaller than the first $\kappa > \omega$ having the tree property, the problem seems to be hopelessy difficult at present.

Note added in proof. Shelah (1990) proved that if $V = L$ and $\kappa = \mathrm{cf}(\kappa) > \omega$ is not weakly compact then there is a graph G on κ with $\chi(G) = \kappa$ such that the chromatic number of every subgraph of size less than κ is $\aleph_0$.

4. Ramsey's theorem and its generalizations to larger cardinals

In this section we are going to investigate *r-partitions of length* γ of an arbitrary set X. That means we are going to consider mappings $f: [X]^r \to \gamma$. Clearly, such a mapping corresponds canonically to an edge disjoint partition $\bigcup_{\alpha<\gamma} H_\alpha$ of $[X]^r$ into r-uniform hypergraphs, indexed by ordinals $\alpha < \gamma$; namely

$$f(e) = \alpha \iff e \in H_\alpha \quad \text{for } \alpha < \gamma \ , \ e \in [X]^r .$$

Since most of the questions we ask depend only on the cardinality of the underlying set, we usually choose X to be a cardinal. The cliques of H_α are called *homogeneous for f in the color* α. It was proved in Ramsey (1930) that:

Theorem 4.1. *For all r partitions f of length $k < \omega$ of an infinite set there is an infinite set homogeneous for f.*

The discovery of this result had an enormous impact on both combinatorics and set theory, as can be seen, e.g., from the two monographs Graham et al. (1980), and Erdős et al. (1984). Referring the reader to these sources, we intend to keep our account of infinite Ramsey theory within limits, concentrating on results we are going to use in other sections of this chapter.

To have a concise notation for the possible generalizations, we define the so-called *ordinary partition symbol*, introduced in Erdős and Rado (1956).

Definition 4.2. $\kappa \rightarrow (\kappa_\alpha)^r_{\alpha<\gamma}$ means that for all r-partitions f of length γ of κ there are α and a set X_α of size κ_α, homogeneous for f in color α.

$\kappa \not\rightarrow (\kappa_\alpha)^r_{\alpha<\gamma}$ denotes the negation of this statement. If all κ_α are equal to λ, we use the notation $\kappa \rightarrow (\lambda)^r_\gamma$.

Other, mostly self-explanatory variants of this notation will be introduced later. The relation in Definition 4.2 remains true if κ is increased or if the variables on the right-hand side are decreased. This was the motivation behind separating them with the "arrow" symbol. Ramsey's theorem states, in this notation, as

$$\omega \rightarrow (\omega)^r_k \quad \text{for } r, k < \omega ,$$

and the Ramsey function $R_r(l_0, \ldots, l_{\kappa-1})$, used in finite combinatorics (see chapter 25), is the smallest integer R, for which

$$R \rightarrow (l_i)^r_{i<\kappa}$$

holds.

The reader is advised to check that the existence of the finite Ramsey function follows from Ramsey's theorem $\omega \rightarrow (\omega)^r_k$, by compactness (i.e., Theorem 3.5).

We now intend to describe a proof of Theorem 4.1, which will be useful for us for other purposes as well. We need some preliminaries. We now assume $r \geqslant 2$.

Let $f: [\kappa]^r \rightarrow \gamma$ be an r-partition of length γ. Note that the underlying set κ has a natural well-ordering. If A, B are subsets of κ, we write $A < B$ if $a < b$ for $a \in A$ and $b \in B$. A subset $X \subseteq \kappa$ is *prehomogeneous for* f if the color of an r-tuple in X does not depend on its last element, i.e., for $e \in [X]^{r-1}$, $\alpha, \beta \in X$ with $e < \{\alpha, \beta\}$ imply that

$$f(e \cup \{\alpha\}) = f(e \cup \{\beta\}) .$$

If $A \subseteq \kappa$, $A < \{\alpha, \beta\} \subset \kappa$, we say that α and β are *twins over* A *for* f if for every $e \in [A]^{r-1}$

$$f(e \cup \{\alpha\}) = f(e \cup \{\beta\}) .$$

We write $\alpha \equiv \beta\ (A, f)$ in this case. Note that if $|A| < r - 1$, $\alpha \equiv \beta\ (A, f)$ always holds for $A < \{\alpha, \beta\}$, and if A and γ are finite, then the number of equivalence classes is finite.

Now we can associate a sequence $(\beta^\alpha_\nu : \nu < \phi_\alpha)$ to each $0 < \alpha < \kappa$ as follows. We define the sequence β^α_ν by recursion on ν. First, $\beta^\alpha_0 = 0$. Assume $(\beta^\alpha_\mu : \mu < \nu)$ is defined for $\mu < \nu$. Set $\{\beta^\alpha_\mu : \mu < \nu\} = B^\alpha_\nu$. Now if there is a β with $B^\alpha_\nu < \{\beta\} < \{\alpha\}$ which is a twin of α over B^α_ν for f, we write β^α_ν for the minimal β of this kind; if not we stop, β^α_ν is not defined, and we set $\phi_\alpha = \nu$. Finally, write $\beta <_f \alpha$ or $\beta < \alpha$, for short, if β occurs in the sequence associated to α, i.e., if $\beta = \beta^\alpha_\nu$ for some $\nu < \phi_\alpha$. After these definitions, it is a matter of easy computation to prove:

Lemma 4.3. *$\beta <_f \alpha$ implies $\beta < \alpha$; $(\kappa, <_f)$ is an S-tree of height $\leqslant \kappa$; $\{\alpha : \phi_\alpha = \nu\}$ is the νth level of this S-tree; every chain of this S-tree is prehomogeneous for f.*

The tree $(\kappa, <_f)$ is called the *canonical partition tree associated with f*.

The use of prehomogeneous sets should be obvious from the following lemma.

Lemma 4.4. *Let $f : [\kappa]^r \to \gamma$. Assume $X \subseteq \kappa$ is prehomogeneous for f, $|X| = \lambda$, X does not have a largest element and $\lambda \to (\kappa_\alpha)^{r-1}_{\alpha<\gamma}$. Then there is a homogeneous set of size κ_α for f in color α, for some $\alpha < \gamma$.*

Proof. The r-partition f induces an $(r-1)$-partition f' of X, by the prehomogeneity. There is a homogeneous set Y for f' of size κ_α in some color α, by the assumption. Y is homogeneous for f as well. □

Now we can easily prove Ramsey's theorem. We use induction on r. For $r = 1, \omega \to (\omega)^1_k$ is Dirichlet's box principle. Assume $r > 1$, $\omega \to (\omega)^{r-1}_k$, and let $f : [\omega]^r \to k$ be an r-partition of length k. By Lemma 4.3, $(\omega, <_f)$ is an S-tree. We want to see that it is an ω S-tree. Since $\phi_n \leq n$ for each $n < \omega$, the height of this tree is ω. We have to show that each level is finite. This is done by induction. Let T_i be the ith level. We have $T_0 = \{0\}$. Assume that $T_0, \ldots, T_i$ are finite. Now, for an $\alpha \in T_{i+1}$, there are only finitely many possibilities for $(\beta^\alpha_j : j \leq i)$, since $\beta^\alpha_j \in T_j$, and among infinitely many α's with the same set of predecessors $B_{i+1} = B^\alpha_{i+1}$, there are two which are twins over B_{i+1} for f. Hence, T_{i+1} must be finite as well. It now follows, from the form 3.3 of König's lemma, that $(\omega, <_f)$ has a branch B. B is clearly infinite and, by Lemma 4.3, it is prehomogeneous for f. Then, by the inductive assumption and by Lemma 4.4, there is an infinite homogeneous set for f. □

We deduced Ramsey's theorem from König's lemma and indeed, the two statements are equally strong. Instead of pondering about the exact logical meaning of this claim, we remark that, indeed, we proved a much stronger result of Erdős and Tarski (1943).

Theorem 4.5. *Assume κ is a strongly inaccessible cardinal which has the tree property. Then*

$$\kappa \to (\kappa)^r_\gamma$$

holds for $\gamma < \kappa$.

Indeed, in the proof of Ramsey's theorem, when establishing that $(\omega, <_f)$ is an ω S-tree, we only used that ω is strongly inaccessible.

The paper by Erdős and Tarski (1943) initiated the modern theory of large cardinals. There it was recognized that the analogue of Ramsey's theorem holds for strongly inaccessible cardinals having the tree property, but fails for cardinals not strongly inaccessible. This results from the following facts:

$$\kappa \not\to (\kappa, \mathrm{cf}(\kappa)^+)^2 \quad \text{and} \quad 2^\kappa \not\to (\kappa^+)^2_2 \quad \text{for } \kappa \geq \omega \;, \tag{4.6}$$

since every cardinal λ not strongly inaccessible is either singular, or for some $\kappa < \lambda$, $2^\kappa \geq \lambda$ holds. Now the first fact stated in (4.6) is quite obvious, while the proof of the second is based on the following idea. Consider ${}^\kappa 2$, i.e., the set of 0, 1 sequences of length κ, order its elements lexicographically, and let $<_{\text{lex}}$ be this ordering. Prove an easy technical lemma, saying that every increasingly or decreasingly well-ordered set in the $<_{\text{lex}}$ ordering, has cardinality at most κ. Also fix a well-ordering $<$ of ${}^\kappa 2$ and define the two graphs G_0 and G_1, where G_0 consists of the pairs parallel in the orderings $<_{\text{lex}}$ and $<$, and G_1 of the rest of the pairs. This is a generalization of a proof of Sierpiński, where the ordering and a well-ordering of $\mathbb{R}$ are compared similarly, and which gave the special case

$$2^{\aleph_0} \not\to (\aleph_1)^2_2 .$$

We turn back to the line of thought of the paper by Erdős and Tarski in the next section. We now prove the Erdős–Rado theorem, which gives the existence of the Ramsey function for infinite cardinals, and even a quantitative bound. First we consider the case $r = 2$.

Theorem 4.7. *For $\kappa \geq \omega$, $(2^\kappa)^+ \to ((2^\kappa)^+, (\kappa^+)_\kappa)^2$, and as an obvious corollary of this, $(2^\kappa)^+ \to (\kappa^+)^2_\kappa$ holds.*

Proof. This is a special case of a more general result, and here we can avoid the use of the canonical partition tree. Set $\lambda = (2^\kappa)^+$. Let $f: [\lambda]^2 \to \kappa$. We assume that there is no homogeneous set of size κ^+ for f in the colors ν, $1 \leq \nu < \kappa$. We will prove that there is a homogeneous set of size λ for f in color 0.

Let $A = \{\alpha < \lambda : \text{cf}(\alpha) = \kappa^+\}$. Considering that $\kappa^+ \leq 2^\kappa < \lambda$, by Fact A.19, this is a stationary subset of λ. For each $\alpha \in A$ and $1 \leq \nu < \kappa$, let $A_{\alpha,\nu}$ be a subset of $\alpha + 1$ $(=\alpha \cup \{\alpha\})$, maximal with respect to the property that it contains α, and it is homogeneous for f in the color ν. Write each $A_{\alpha,\nu}$ in the form $B_{\alpha,\nu} \cup \{\alpha\}$, where $B_{\alpha,\nu} \subseteq \alpha$. Let $B_\alpha = \bigcup_{1 \leq \nu < \kappa} B_{\alpha,\nu}$. By assumption, $|B_{\alpha,\nu}| \leq \kappa$, and hence $|B_\alpha| \leq \kappa$. Let $g(\alpha) = \sup B_\alpha$, for $\alpha \in A$. Then, by $\text{cf}(\alpha) = \kappa^+$, $g(\alpha) < \alpha$ for $\alpha \in A$, hence g is regressive on A. By Theorem A.18, there is a stationary subset $A_0 \subseteq A$ and an ordinal $\xi < \lambda$, such that $g(\alpha) = \xi$ for $\alpha \in A_0$. Considering that $\xi < \lambda$, we have $|\xi| \leq 2^k$, and so $|\xi|^\kappa \leq 2^{\kappa^2} \leq 2^\kappa < \lambda$. Therefore, there is a stationary subset $A_1 \subseteq A_0$ such that for $\alpha, \beta \in A_1$ and for all $1 \leq \nu < \kappa$, $B_{\alpha,\nu} = B_{\beta,\nu}$ holds. We claim that A_1 is homogeneous for f in color 0. Indeed, if $f(\{\alpha, \beta\}) = \nu > 0$ for a pair $\alpha, \beta \in A_1$, and say $\beta < \alpha$, then $\alpha + 1 \supseteq B_{\alpha,\nu} \cup \{\beta, \alpha\} \supset B_{\alpha,\nu} \cup \{\alpha\}$, and $B_{\alpha,\nu} \cup \{\beta, \alpha\}$ is homogeneous for f in color ν, a contradiction. □

Now we give the Erdős–Rado theorem for r-uniform hypergraphs. Let $\exp_0(\kappa) = \kappa$, and $\exp_{r+1}(\kappa) = 2^{\exp_r(\kappa)}$, for $r < \omega$. Then

Theorem 4.8. $(\exp_{r-1}(\kappa))^+ \to (\kappa^+)^r_\kappa$ *for $r \geq 1$, $\kappa \geq \omega$.*

Proof. For $r = 1$, this is $\kappa^+ \to (\kappa)^1_\kappa$, which is true, since κ^+ is a regular cardinal.

We use induction. Assume the statement is true for r. To prove the statement for $r+1$, by Lemma 4.4, it is sufficient to prove the following stepping-up lemma.

Lemma 4.9. *Let $\lambda \geq \omega$, and let $f : [(2^\lambda)^+]^{r+1} \to 2^\lambda$. Then there is a subset $X \subset (2^\lambda)^+$ of size λ^+ such that f is prehomogeneous on X.*

Proof of Lemma 4.9. Again, this is not the strongest possible statement, so we can avoid the use of the canonical partition tree. Set $\rho = (2^\lambda)^+$, and let $f : [\rho]^{r+1} \to 2^\lambda$. Again let $A = \{\alpha < \rho : \mathrm{cf}(\alpha) = \lambda^+\}$. Just as in the proof of Theorem 4.7, A is a stationary subset of ρ. For each $\alpha \in A$, let A_α be a maximal subset of $\alpha + 1$ containing α and prehomogeneous for f. A_α is of the form $B_\alpha \cup \{\alpha\}$, where $B_\alpha \subseteq \alpha$. If $|B_\alpha| \geq \lambda^+$ for some $\alpha \in A$, we are done. So we may assume instead that $|B_\alpha| \leq \lambda$ for each $\alpha \in A$. Put $g(\alpha) = \sup B_\alpha$. Then $g(\alpha) < \alpha$ for $\alpha \in A$, since $\mathrm{cf}(\alpha) = \lambda^+$. It follows from Theorem A.18 that there are $\xi < \lambda$ and an $A_0 \subseteq A$, $|A_0| = \rho$, such that $g(\alpha) = \xi$ for $\alpha \in A_0$. Since $\xi < \rho$ implies $|\xi|^\lambda \leq 2^\lambda < \rho$, there are $B \supset \xi$ and an $A_1 \subseteq A_0$, $|A_1| = \rho$, such that $B_\alpha = B$ for $\alpha \in A_1$. The number of equivalence classes for $\alpha \equiv \beta$ (B, f), is at most $(2^\lambda)^{|B|^{r-1}} = 2^\lambda$. Since $|A_1| = \rho > 2^\lambda$, there are $\alpha \neq \beta \in A_1$ such that α and β are twins for f over B. If say $\beta < \alpha$, then $B \cup \{\beta, \alpha\}$ is still a subset of $\alpha + 1$ prehomogeneous for f, a contradiction. □

Note that Theorem 4.7 is stronger than the instance $r = 2$ of Lemma 4.9. The last aim of this section is to convince the reader that the seemingly enormous bound given by Theorem 4.8 is best possible. The following is a result of Erdős et al. (1965).

Theorem 4.10. $\exp_{r-1}(\kappa) \not\to (\kappa^+)^r_2$ *for $\kappa \geq \omega$ and $r \geq 2$.*

For $r = 2$, this is $2^\kappa \not\to (\kappa^+)^2_2$. We stated this in (4.6). I will only outline the proof for $r = 3$, i.e., the proof of $2^{2^\kappa} \not\to (\kappa^+)^3_2$. Let $\lambda = 2^\kappa$, and $S = {}^\lambda 2$, the set of 0, 1 sequences of length λ. For $f \neq g \in S$, let $\delta(f, g) = \min\{\alpha < \lambda : f(\alpha) \neq g(\alpha)\}$, the so-called first discrepancy of f and g. The lexicographical ordering $<_{\mathrm{lex}}$ of S is defined by $f <_{\mathrm{lex}} g \Leftrightarrow f(\delta(f, g)) = 0$. Let $<$ be a well-ordering of S. Let $\phi : [\lambda]^2 \to 2$ be a 2-partition of $\lambda = 2^\kappa$ establishing $\lambda \not\to (\kappa^+)^2_2$. Let $F = \{f_0, f_1, f_2\}$, $f_0 < f_1 < f_2$, be an arbitrary triple of S. First, we split the set of the triples of S into four parts:

$$K_0 = \{F : f_0 <_{\mathrm{lex}} f_1 \,{}_{\mathrm{lex}}\!> \mathrm{f}_2\}\,, \quad K_1 = \{F : f_0 \,{}_{\mathrm{lex}}\!> f_1 <_{\mathrm{lex}} f_2\}\,,$$
$$K_2 = \{F : f_0 <_{\mathrm{lex}} f_1 <_{\mathrm{lex}} f_2\}\,, \quad K_3 = \{F : f_0 \,{}_{\mathrm{lex}}\!> f_1 \,{}_{\mathrm{lex}}\!> f_2\}\,.$$

Let $K = K_2 \cup K_3$ be the class of "smooth" triples. For $F \in K$, let $\Delta(F) = \{\delta(f_0, f_1), \delta(f_1, f_2)\}$. It is easy to see that $\Delta(F)$ has two elements for $F \in K$, and so $\Delta(F) \in [\lambda]^2$. Now we can define a partition $\Psi : [S]^3 \to 2$. Let $\Psi(F) = 0$ for $F \in K_0$, and $\Psi(F) = 1$ for $F \in K_1$. For $F \in K$ put $\Psi(F) = \phi(\Delta(F))$.

To see that this works, we need two sublemmas.

Lemma. (i) *Assume that for some $X \subseteq S$ and $\tau \geqslant \omega$, $|X| = \tau$ and $[X]^3 \cap K_i = \emptyset$ for some $i < 2$. Then there is a $Y \subseteq X$, $|Y| = \tau$, such that $[Y]^3 \subset K_2$ or $[Y]^3 \subset K_3$.*

(ii) *Assume $X \subseteq S$, $|X| = \tau$ is regular, and $<$ coincides with $<_{\text{lex}}$ or ${}_{\text{lex}}>$ on X. Then there is a sequence $\{f_\nu : \nu < \tau\} \subseteq X$, and a sequence $\{\delta_\nu : \nu < \tau\} \subseteq \lambda$, such that $\delta(f_\mu, f_\nu) = \delta_\mu$ for all $\mu < \nu < \tau$.*

I leave the proofs to the reader. Assume now that $X \subseteq S$ is homogeneous for Ψ, say, in class $i < 2$, and $|X| = \kappa^+$. By (i) we may assume that $<_{\text{lex}}$ or ${}_{\text{lex}}>$ coincides with $<$ on X; hence all triples of X are from K. We may assume that $X = \{f_\nu : \nu < \kappa^+\}$ is increasing in the well-ordering, and satisfies (ii) with a sequence $\{\delta_\nu : \nu < \kappa^+\}$. Let $\mu < \nu$. Then $\{\delta_\mu, \delta_\nu\} = \Delta(\{f_\mu, f_\nu, f_{\nu+1}\})$, and hence $\phi(\{\delta_\mu, \delta_\nu\}) = i$, and the set $\{\delta_\nu : \nu < \kappa^+\}$ is homogeneous for ϕ in the color $i < 2$, a contradiction to the choice of ϕ. □

5. Combinatorics and the discovery of large cardinals

We give the definition of inaccessible cardinals in the Appendix. Let us define now the so-called cumulative hierarchy of sets. First, $V_0 = \emptyset$. Then, in general, $V_{\alpha+1}$ = the set of all subsets of V_α, $V_\alpha = \bigcup_{\beta<\alpha} V_\beta$ for a limit ordinal α. It was J. von Neumann, who gave the first exact proof that we may assume (it is relatively consistent) that the class of all sets V consists of $\bigcup_{\alpha \text{ an ordinal}} V_\alpha$, but the fact that the sets V_κ, for $\kappa > \omega$ strongly inaccessible, are natural models for the axioms of set theory, was already known to Zermelo. The assumption that such cardinals exist seemed to be a natural extension of the axiom-system, working in the way to capture "Cantor's absolute", whatever that might mean. But even with a more modest and realistic approach if, as it was clear, set theory is not sufficient to decide within it all problems of mathematics, any natural strengthening of the axiom-system should be welcome. Quite early in the game, Mahlo (1911) suggested further large cardinal assumptions. To understand these let us introduce some strengthenings step by step. First we may assume that for every λ there is a $\kappa > \lambda$ that is strongly inaccessible. Then we can arrange these inaccessible cardinals in an increasing sequence κ^1_α, and wonder if there is a "fixed point", i.e., if there is one with $\kappa^1_\alpha = \alpha$? Now assume first there is one, and then assume there is one greater than λ for every λ. Denote the increasing sequence of those by κ^2_α, assume there is one with $\kappa^2_\alpha = \alpha$, and iterate this transfinitely. The observation of Mahlo, in present day terminology, was that none of these yield the existence of a κ such that

$$\begin{aligned}&\kappa \text{ is strongly inaccessible, and the set of cardinals } \lambda < \kappa\\ &\text{which are strongly inaccessible is stationary in } \kappa.\end{aligned} \tag{5.1}$$

He suggested assuming that cardinals satisfying (5.1) exist. These are nowadays called Mahlo cardinals, or 1-Mahlo cardinals. Of course, there is an immediate possibility to iterate even this transfinitely, e.g., κ is a 2-Mahlo cardinal if $\{\lambda < \kappa : \lambda \text{ is Mahlo}\}$ is stationary in κ, and so on.

Though this approach was not particularly fruitful in 1911, Mahlo cardinals play an important role in present day set theory. Let us remark that, similarly to the case of inaccessible cardinals, it is very easy to show the consistency that they *do not* exist. Assuming that there is a κ, for which, say, (5.1) holds, then there is a smallest one, say κ_0, and it is easy to see that V_{κ_0} is a model, in which there is no κ satisfying (5.1). The situation is similar with all the large cardinal assumptions considered in this section.

Another problem, investigated by the Polish set theory school in the 1930s, was the measure problem, raised by Banach in Banach and Kuratowski (1929). For what cardinals κ does there exist a probability measure on all subsets of κ, vanishing on one-point sets? Ulam (1930) proved that any such cardinal must be weakly inaccessible. Soon it was discovered that the core of this problem is to decide the question in case when the measure can take only the values 0 and 1, and Tarski proved that a cardinal κ carrying such a measure must be strongly inaccessible. In the paper by Erdős and Tarski (1943) where they investigated a problem in the theory of partially ordered sets, they obtained an answer involving inaccessible cardinals, and in the remarks they formulated the basic connections about different properties of large cardinals.

Definition 5.2. An infinite cardinal κ is *measurable*, if there is a free ultrafilter U on the subsets of κ which is κ-complete, i.e., the intersection of fewer than κ sets in U is in U.

Note that, by Theorem A.14, ω is measurable. By defining

$$\mu(X) = 1 \Leftrightarrow X \in U ,$$

we see how an ultrafilter corresponds to a 0–1 measure defined on all subsets of κ.

Erdős and Tarski proved that if a strongly inaccessible κ has the tree property, then $\kappa \rightarrow (\kappa)^r_\gamma$ holds for $r < \omega$ and $\gamma < \omega$. We proved this in Theorem 4.5, and they proved:

Theorem 5.3. *If κ is measurable, then κ has the tree property.*

Proof. Let $(T, <)$ be a κ S-tree. We know that $|T| = \kappa$. Let U be an ultrafilter on T satisfying the condition in Definition 5.2. For each $x \in T$, let $\check{x} = \{y \in T : x \leqslant y\}$. Clearly, for each $\alpha < \kappa$, $T = \bigcup_{\beta < \alpha} T_\beta \cup \bigcup \{\check{x} : x \in T_\alpha\}$. Since $|T_\alpha| < \kappa$ for each $\alpha < \kappa$, there is exactly one $x_\alpha \in T_\alpha$ with $\check{x}_\alpha \in U$. Since $\check{x}_\alpha \cap \check{x}_\beta \in U$ for $\alpha, \beta < \kappa, x_\alpha$ and x_β have a common upper bound, and therefore x_α and x_β are comparable in T; hence $\{x_\alpha : \alpha < \kappa\}$ is a branch of T. □

Unfortunately, Erdős and Tarski guessed wrong. They speculated that it might be consistent with the axioms of set theory that all strongly inaccessible cardinals are measurable. It took eighteen more years to discover that this is not so. In 1960, Hanf, then a student of Tarski, proved that a compactness property, similar to that formulated in Theorem 3.4 for ω, does not hold for κ, if κ is the first strongly inaccessible cardinal. He used languages with infinitely long formulas, and proved that a weak form of Gödels compactness theorem does not hold for these languages. He also proved (see Hanf 1964) that the Ramsey property is really equivalent to the König property, i.e.,

Theorem 5.4. *If for some $\kappa \geqslant \omega$ $\kappa \rightarrow (\kappa)^2_2$ holds, then κ has the tree property.*

We postpone the proof, and first we finish the historical remarks. After Theorem 5.4 was proved, strongly inaccessible cardinals having the tree property were called *weakly compact cardinals*, and their class is called C_0, $C_1 \subseteq C_0$ is the class of measurable cardinals, $C_2 \subseteq C_1$ is the class of strongly compact cardinals (see our definition after Theorem 3.6). In Kiesler and Tarski (1964) an attempt is made, to transform the logic proofs to a purely set-theoretic proof of the fact that cardinals in C_0 are very large, e.g., $\kappa \in C_0$ is α-Mahlo for all $\alpha < \kappa$. However, all known proofs bear some trace of logic in them.

It is easier to devise direct combinatorial proofs that the first strongly inaccessible cardinal greater than ω, and the cardinals in the small Mahlo classes, are not weakly compact. After all, if there is a graph G on κ, the first strongly inaccessible cardinal $>\omega$, with $\alpha(G)$ and $\omega(G) < \kappa$, we would like to see one. We will try to do our best in Theorem 5.6. First we prove Theorem 5.4.

Proof of Theorem 5.4. Assume $(T, \prec)$ is a κ S-tree; let $T_\alpha : \alpha < \kappa$ denote its levels. We now describe a procedure, called the *squashing of the tree*, which extends the partial order $<$ of T to a total order $<^*$. For $x \in T$ and $\beta \leqslant \alpha(x)$, let x_β be the unique element y of T with $y \in T_\beta$ and $y \leqslant x$. For $x, y \in T$, let $\delta(x, y) = \min\{\beta : x_\beta \neq y_\beta\}$ if x and y are incomparable. $<^*$ is very similar to a lexicographic order. For each $\alpha < \kappa$ choose an arbitrary ordering $<_\alpha$ of the level T_α, and put

$$x \leqslant {}^*y \;\Leftrightarrow\; \text{either } x \leqslant y \quad\text{or}\quad x \text{ and } y \text{ are incomparable and}$$
$$\text{for } \alpha = \delta(x, y)\;\; x_\alpha \leqslant_\alpha y_\alpha\,.$$

It is easy to check that $\leqslant^*$ is a total order of T, and that

$$x <^* y \text{ implies that}$$
$$x_\alpha \leqslant^* y_\alpha \quad \text{for } \alpha(x), \alpha(y) \geqslant \alpha\,.$$

Now let $<_0$ be a well-ordering of T. Using the assumption $\kappa \rightarrow (\kappa)^2_2$, it now follows from the idea of Sierpiński, described in the proof of (4.6), that there exists a set $S \subseteq T$, $|S| = T$, which is either increasingly or decreasingly well-ordered in the ordering $<^*$. We may assume that $(S, <^*)$ is well-ordered and of order-type κ. We know from (4.6) that κ is regular. It follows now from (5.5) that there is an element x^α of T_α such that $y = x^\alpha$ holds, for all but fewer than κ elements y of S. Then $\{x^\alpha : \alpha < \kappa\}$ is a branch of $(T, <)$. □

We now prove:

Theorem 5.6. *Let $\kappa > \omega$ be the first strongly inaccessible cardinal. Then $\kappa \not\to (\kappa)^2_2$.*

Proof. By Theorem 5.4 we only have to prove that κ does not have the tree property. Let B denote the set of infinite strong limit cardinals $<\kappa$, i.e., $B = \{\lambda < \kappa : \lambda$ is a limit cardinal, and for all $\tau < \lambda$, $2^\tau < \lambda$ holds$\}$. It is very easy to see, using that κ is strongly inaccessible, that B is a club set in κ. We will need the following sublemma.

Lemma 5.6′. *For $\lambda \in B$ there is a one-to-one regressive function on $B \cap \lambda$.*

Using this lemma we can easily finish the proof. Let T be the set of functions f, such that $\mathrm{Dom}(f) = \lambda \cap B$ for some $\lambda \in B$, and f is one-to-one and regressive on $\lambda \cap B$. Define $<$ on T, by $f < g \Leftrightarrow f \subset g$. $(T, <)$ is clearly an S-tree. We can see what the αth level of this tree is. If $\{\lambda_\alpha : \alpha < \kappa\}$ is an increasing enumeration of the elements of B, then $T_\alpha = \{f \in T : D(f) = \lambda_\alpha \cap B\}$. It is also clear that $|T_\alpha| \leqslant \lambda_\alpha^{\lambda_\alpha} = 2^{\lambda_\alpha^2} = 2^{\lambda_\alpha} < \kappa$, for $\alpha < \kappa$. The sublemma 5.6′ tells us that $T_\alpha \neq \emptyset$ for $\alpha < \kappa$, and hence the height of T is κ and $(T, <)$ is a κ S-tree. On the other hand, $(T, <)$ has no branch, since if B is a branch, then $F = \bigcup\{f : f \in B\}$ would be a one-to-one regressive function on B. By Theorem A.18 this contradicts the fact that B is stationary. Hence κ does not have the tree property. □

Before proving Lemma 5.6′, we prove the following:

Lemma 5.6″. *Assume $\tau \in B$, f_τ is a one-to-one regressive function on $B \cap \tau$, and $\rho < \tau$ is a cardinal. Then there is a function f'_τ having the same property, and such that, for $\rho^+ \leqslant \sigma \in B$, $f'_\tau(\sigma) < \rho^+$ implies that $f'_\tau(\sigma) > \rho$ and $\{f'_\tau(\sigma) : \sigma \in B \cap [\rho^+, \tau)\}$ omits a subset of size ρ^+ from (ρ, ρ^+).*

Proof. Indeed, $|(\rho, \rho^+)| = \rho^+$, by Theorem A.7, and f_τ being one-to-one, implies there are at most ρ^+ among the $\sigma \geqslant \rho$ in B with $f_\tau(\sigma) < \rho^+$. We only have to change the values of f_τ for these σ, and this can obviously be done. □

Proof of Lemma 5.6′. By transfinite induction on λ. Lemma 5.6′ is obvious for the minimal element of B which is ω. Assume $\lambda > \omega$, and that there is an f_τ: regressive and one-to-one on $B \cap \tau$ for every $\tau < \lambda$, $\tau \in B$. Now, if λ is not a limit point of B, then B being closed, implies $B \cap \lambda$ has a largest element τ. By Lemma 5.6″, we change f_τ so that its range omits one ordinal $\xi < \tau$, we define $f_\lambda(T) = \xi$ and $f_\lambda(\sigma) = f_\tau(\sigma)$ for $\sigma \in \tau \cap B$. Hence we may assume $\lambda = \sup B \cap \lambda$. Since κ is the first inaccessible $>\omega$, λ is singular, $\mathrm{cf}(\lambda) < \lambda$. There is an increasing continuous sequence $\{\lambda_\xi : \xi < \mathrm{cf}(\lambda)\}$ of cardinals $<\lambda$ with $\sup\{\lambda_\xi : \xi < \mathrm{cf}(\lambda)\} = \lambda$ and $\lambda_0 = \mathrm{cf}\{\lambda)$. Continuity means that $\lambda_\xi = \sup\{\lambda_\eta : \eta < \xi\}$ for $\xi < \mathrm{cf}(\lambda)$ if ξ is a limit ordinal. By continuity $\lambda = \lambda_0 \cup \bigcup_{\xi < \mathrm{cf}(\kappa)} [\lambda_\xi, \lambda_{\xi+1})$. Define, by Lemma 5.6″, $f_{\lambda_{\xi+1}}$

for $\xi < \lambda_0$ so that $\lambda_\xi < f_{\lambda_{\xi+1}}(\tau)$ for $\lambda_\xi^+ \leqslant \tau < \lambda_{\xi+1}$, and $(\lambda_1 \setminus \lambda_0) \setminus \mathrm{Ran}(f_{\lambda_1})$ has cardinality λ_0^+, and $\lambda_0 \setminus \mathrm{Ran}(f_{\lambda_0})$ is non-empty. Define $f_\lambda(\tau) = f_{\lambda_0}(\tau)$ for $\tau < \lambda_0$, $f_\lambda(\lambda_0) \in \lambda_0 \setminus \mathrm{Ran}(f_{\lambda_0})$, $f_\lambda(\tau) = f_{\lambda_{\xi+1}}(\tau)$ for $\lambda_\xi < \tau < \lambda_{\xi+1}$, and extend f_λ to $\{\lambda_\xi : 1 \leqslant \xi < \mathrm{cf}(\lambda)\}$ using a one-to-one function mapping this set into $(\lambda_1 \setminus \lambda_0) \setminus \mathrm{Ran}(f_{\lambda_1})$. □

The proof of Theorem 5.6 extends to all κ, for which there are sets of cardinals $<\kappa$, $A \subseteq B$, A stationary in κ, B closed in κ, and for all $\lambda \in B$, λ regular, $\lambda \cap A$ is nonstationary in λ. This is easily seen for cardinals κ in the first ω levels of the Mahlo hierarchy. The proof, of course, comes from general reflection principles, valid for weakly compact cardinals.

To take up the thread of the tale again, large cardinals were also approached through another combinatorial problem. In Erdős and Rado (1956) the following problem was considered. Let $f_r : [\kappa]^r \to \gamma$ be a sequence of r-partitions of length γ of κ, for $r \in \mathbb{N}$. For what cardinals λ does there exist a set $X \subset \kappa$, $|X| = \lambda$, which is simultaneously homogeneous for all these partitions. They briefly denoted this statement by $\kappa \to (\lambda)_\gamma^{<\omega}$, and the negation of it by $\kappa \not\to (\lambda)_\gamma^{<\omega}$. They proved $2^\omega \not\to (\omega)_2^{<\omega}$ for the first attempt. Later, in Erdős and Hajnal (1958) it was proved that $\lambda \not\to (\omega)_2^{<\omega}$ holds for all λ less than the first uncountable strongly inaccessible cardinal, and more importantly, it was proved that:

Theorem 5.7. *For $\kappa > \omega$ measurable, $\kappa \to (\kappa)_\gamma^{<\omega}$ holds for $\gamma < \kappa$.*

We intend to prove this theorem soon, but first we make some remarks. $\kappa \to (\kappa)_2^{<\omega}$ is the same as $\kappa \to (\kappa)_\gamma^{<\omega}$ for every $\gamma < \kappa$. Cardinals with this property are called Ramsey cardinals; cardinals with the weaker property $\kappa \to (\omega_1)_2^{<\omega}$ are called Erdős cardinals. One important property of the measurable and generally the large cardinals is that they change the small sets of the universe as well. The set of reals will have different concrete mathematical properties, if we assume the existence of different large cardinals. The earliest results of this type were proved through these combinatorial properties. It was proved in Scott (1961) that the existence of a measurable cardinal contradicts Gödel's axiom of constructibility. Rowbottom (1971) proved that the existence of an Erdős cardinal implies that there are only countably many constructible reals, and Solovay (1969) proved that the existence of an Erdős cardinal implies that every Π_1^1 set of reals (analytic complement) is either countable or has a perfect subset. The deepest results for $\kappa \to (\lambda)_2^{<\omega}$ were proved in Silver (1966). Here it is proved that the first cardinal κ with $\kappa \to (\omega)_2^{<\omega}$ is very large, in fact, larger than the first weakly compact cardinal, but it still can exist in the constructible universe. Also much stronger conclusions are drawn from the existence of an Erdős cardinal than the ones mentioned before. However, lacking space we cannot go into all that.

For the proof of Theorem 5.7 we will need a concept introduced by Scott.

Let $\kappa > \omega$ be a regular cardinal, and let U be an ultrafilter on the subsets of κ. U is said to be *normal* if it is free and for every $A \in U$ and for every regressive function f on A, f is constant almost everywhere, i.e., there are a $B \subseteq A$, $B \in U$, and $\xi < \kappa$, such that $f(\alpha) = \xi$ for $\alpha \in B$.

Lemma 5.8. *If $\kappa > \omega$ is measurable, then κ carries a κ-complete normal ultrafilter.*

Proof (*Outline*). Let V be a κ-complete free ultrafilter on κ. Let $S = {}^{\kappa}\kappa$. For $f, g \in S$, define $f < g \Leftrightarrow \{\xi < \kappa : f(\xi) < g(\xi)\} \in V$. There is no decreasing sequence $f_0 > \cdots > f_n > \cdots$. Indeed, otherwise, for $n < \omega$,

$$A_n = \{\xi < \kappa : f_{n+1}(\xi) < f_n(\xi)\} \in V \,.$$

Hence $\bigcap_{n<\omega} A_n \in V$, but then, for some $\xi \in \bigcap_{n<\omega} A_n$, $f_0(\xi) > \cdots > f_n(\xi) > \cdots$ would be a decreasing sequence of ordinals. We write $f \equiv g\ (V)$ in case $\{\xi < \kappa : f(\xi) = g(\xi)\} \in V$, and $\bar{\xi}$ for the constant ξ function.

Let $S^* = \{f \in S : f \not\equiv \bar{\xi}$ for any $\xi < \kappa\}$. Then S^* is non-empty, as is shown by the existence of the identity function. By the well-foundedness of $<$, S^* has a $<$-minimal element f_0. This means that if $g(\xi) < f_0(\xi)$ on a set in U, then $g \equiv \bar{\eta}$ for some $\eta < \kappa$. Now let $U = \{A \subset \kappa : f_0^{-1}(A) \in V\}$. It is a matter of easy computation to see that U is a κ-complete normal ultrafilter on κ.

We next need a lemma of Rowbottom (1971).

Lemma 5.9. *Assume U is a κ-complete normal ultrafilter on $\kappa, r \in \mathbb{N}$, and let $f : [\kappa]^r \to \gamma$ be an r-partition of length γ of κ for some $\gamma < \kappa$. Then there is a set $X \in U$ homogeneous for f.*

Proof. By induction on r. For $r = 1$ the statement follows simply from the κ-completeness of U. Assume the statement is true for $r \geq 1$, and let $f : [\kappa]^{r+1} \to \gamma$ be an $(r+1)$-partition of length γ of κ. Let $f_\alpha : [\kappa - (\alpha + 1)]^r \to \gamma$ be defined by:

$$f_\alpha(e) = f(\{\alpha\} \cup e) \quad \text{for } e \in [\kappa]^r \,,\ \{\alpha\} < e \,.$$

By the induction hypothesis, there is an ordinal ν_α and a set $Y_\alpha \in U$ homogeneous for f_α in the color ν_α for $\alpha < \kappa$.

Let $Y = \{\beta < \kappa : \beta \in \bigcap_{\alpha<\beta} Y_\alpha\}$. We claim that $Y \in U$. Otherwise $\kappa \backslash Y = Z \in U$. For $\beta \in Z$ we can define a $g(\beta) < \beta$ so that $\beta \notin Y_{g(\beta)}$. But then $g(\beta) = \alpha$ for some $\alpha < \kappa$ on a subset $Z' \subseteq Z$, $Z' \in U$, and hence, $Z' \cap Y_\alpha = \emptyset$, a contradiction. Now there is a $W \subseteq Z$, $W \in U$, and $\nu < \gamma$ such that $\nu_\alpha = \nu$ for $\alpha \in W$. Clearly, W is homogeneous for f in color ν. □

Lemmas 5.8 and 5.9 easily yield Rowbottom's proof of the Erdős–Hajnal theorem 5.7.

Proof of Theorem 5.7. Let $\kappa > \omega$ be measurable, $\gamma < \kappa$, and let $f_r : [\kappa]^r \to \gamma$ be given for $r \in \mathbb{N}$. By Lemma 5.8, we can choose a κ-complete normal ultrafilter U on κ. By Lemma 5.9, for each $r \in \mathbb{N}$, there is a $Y_r \in U$ homogeneous for f_r. Then $Y = \bigcap_{r \in \mathbb{N}} Y_r \in U$ is homogeneous for each $r \in \mathbb{N}$. □

Before finishing this section there is one final point I want to make. First I want to illustrate it with an example. I formulate a very innocent looking problem of Erdős. Two subsets of ω_1 are called almost disjoint if their intersection is countable. It is one of the most elementary theorems of set theory that there is a family of size larger than ω_1 of pairwise almost disjoint subsets of ω_1.

Problem 5.10. Does there exist a family of size larger than ω_1 of pairwise almost disjoint *stationary* subsets of ω_1?

The following results are known. It is not hard to see that an affirmative answer is consistent. A yes answer holds in the constructible universe. The consistency of a no answer *implies* the consistency of a measurable cardinal, and is implied by the consistency of a supercompact cardinal (this is something even "larger" than strongly compact). The exact consistency strength of Problem 5.10 is not known (see Foreman et al. 1988).

As is shown by this example, an infinite combinatorial problem, not involving large cardinals, may be untreatable without them. The moral of this is that there is no infinite combinatorics without large cardinals.

6. The coloring number and the chromatic number of infinite graphs

As we have already mentioned, we only consider simple graphs. The concept of the *coloring number* of a graph was introduced in Erdős and Hajnal (1966). Let $<$ be an ordering of V. $<$ is a κ*-ordering* for G if for every vertex $x \in V$ the set $N(<, x)$, i.e., the set of neighbors of x smaller than x, has cardinality less than κ. The *coloring number* of G, denoted by $\mathrm{Col}(G)$, is the smallest cardinal κ for which there exists a κ-well-ordering of G. Clearly, $\mathrm{Col}(G) \leqslant \hat{\Delta}(G)$, and it is also obvious that here inequality can hold. Hence the following is a strengthening of Theorem 2.1.

Theorem 6.1. $\chi(G) \leqslant \mathrm{Col}(G)$ *for every graph* G.

However, the proof of Theorem 2.1 we described gives Theorem 6.1 as well. The example of $\kappa_{\lambda,\lambda}$ shows that the coloring number can be arbitrarily large, while the chromatic number is two.

Note added in proof. This concept was later introduced in finite combinatorics under a different name. In Bollobás (1978) G is called k-degenerate iff $\mathrm{Col}(G) \leqslant k + 1$.

The following is another nice fact to know about the coloring number.

Theorem 6.2. *Assume* $\mathrm{Col}(G) = \kappa$. *Then* G *has a* κ*-well-ordering* $<$ *with* $\mathrm{typ}\, V(<) = |G|$, *the smallest possible type.*

I leave the proof to the reader.

A well-known result of Erdős (1959) (see chapter 4) tells us that for every $r \in \omega$ there is a finite graph G, with chromatic number and girth larger than r; and as a corollary of this, for every $r \in \omega$ there is a countable graph, with girth larger than r. The story for uncountable chromatic number is quite different. We will show that a graph of uncountable chromatic number must contain every even circuit. On the other hand, let us define the *odd girth* of a graph as the length of the smallest odd circuit contained in it. We will show that, for every $\kappa \geqslant \omega$ and $r < \omega$, there are graphs $\chi(G) = \kappa$ and odd girth greater than r, even such that $|V| = \kappa$. We will also prove that for a graph of chromatic number $> \omega$, there is an $r < \omega$ such that it contains all circuits of length $> r$. For the history of these results see the paper by Erdős and Hajnal already mentioned.

More generally, it was proved in the above paper that a graph of uncountable chromatic number must contain a K_{r,ω_1} for every $r < \omega$ and hence, every finite bipartite graph.

We are going to prove an even stronger result of Hajnal and Komjáth (1984). For this we have to define (the isomorphism type of) certain special graphs. $H_{\lambda,\lambda}$ is *the* λ *by* λ *half-bipartite graph*. The vertex set of $H_{\lambda,\lambda}$ is the disjoint union of two well-ordered sets $\{x_\alpha : \alpha < \lambda\}$ and $\{y_\alpha : \alpha < \lambda\}$, and the edges of $H_{\lambda,\lambda}$ are the pairs $\{x_\alpha, y_\beta\}$, for $\alpha < \beta$. Note that the vertices x_α have degree λ while the vertices y_α have degree $< \lambda$. The graph $H_{\lambda,\lambda} + 1$ consists of $H_{\lambda,\lambda}$ with one added point y adjacent to all x_α; $H_{\lambda,\lambda} + 2$ is $H_{\lambda,\lambda}$ with two added points y and z both adjacent to all x_α. $H_{\lambda,\lambda} + \Phi$ is $H_{\lambda,\lambda}$ with λ^+ points C added , such that there are $y \in C$ and a sequence $C = C_0 \supseteq \cdots \supseteq C_\alpha \supseteq \cdots$ for $\alpha < \lambda$ with $\bigcap_{\alpha<\lambda} C_\alpha \supset \{y\}$, $|C_\alpha| > \lambda$ and $C_\alpha \subset N(x_\alpha)$ for $\alpha < \lambda$. Hence $H_{\omega,\omega} + \Phi$ contains $H_{\omega,\omega} + 1$ and K_{r,ω_1} for $r < \omega$.

Theorem 6.3. *Assume* $\mathrm{Col}(G) > \omega$. *Then* G *contains* $H_{\omega,\omega} + \Phi$.

Proof. We prove the statement by transfinite induction on the cardinality of G. The theorem is obviously true for all G with $|V| \leqslant \omega$. Assume $|G| = \kappa > \omega$, and the statement is true for all graphs of size less than κ. We also assume that $H_{\omega,\omega} + \Phi \not\subseteq G$, and we will prove $\mathrm{Col}(G) \leqslant \omega$. For an arbitrary finite subset $X \subset V$, we will denote by $\mathrm{CN}(X)$ the set of common neighbors of X, i.e., the set

$$\bigcap \{N(v) : v \in X\} .$$

Now we define $F : [V]^{<\omega} \to [V]^{<\omega}$ as follows. For $X \in [V]^{<\omega}$, let $F(X) = \mathrm{CN}(X)$ in case $\mathrm{CN}(X)$ is countable, and $F(X) = \emptyset$ otherwise. We will say that a set $A \subseteq V$ is closed if it is closed with respect to F, i.e., $X \in [A]^{<\omega}$ implies $F(X) \subseteq A$. It is easy to see that for each $B \subset V$ there is a smallest closed set containing B. We will denote this by $\bar{B}$. State that $|B| = |\bar{B}|$ holds for any infinite B. Now again a standard argument gives that for $V = \kappa$ there is an increasing continuous sequence $\{A_\alpha : \alpha < \kappa\}$ of closed subsets of V such that $|A_\alpha| < \kappa$ for $\alpha < \kappa$. Indeed, this sequence can be defined by transfinite recursion so that $A_{\alpha+1}$ is always a closed extension of A_α, and $A_\beta = \bigcup_{\alpha<\beta} A_\alpha$ for every limit ordinal β. Here we use the fact that the union of an increasing sequence of closed sets is closed. To exhaust V, one has to fix a well-ordering $<$ of type κ of V, choose the $<$-minimal element x_α of $V \setminus A_\alpha$, and set $A_{\alpha+1} = \overline{A_\alpha \cup \{x_\alpha\}}$. Choosing $A_0 = \emptyset$, it is easy to prove by induction that $|A_\alpha| \leqslant |\alpha| + \omega < \kappa$. Let $B_\alpha = A_{\alpha+1} \setminus A_\alpha$. By the continuity of the A_α sequence, the "rings" B_α are disjoint and their union is V. We now claim that $|N(y) \cap A_\alpha| < \omega$ for $y \in B_\alpha$, $\alpha < \kappa$. Assume indirectly, that for some $\alpha < \kappa$ and $y \in B_\alpha$, $N(y) \cap A$ is infinite and let $\{x_n : n < \omega\}$ be a one-to-one enumeration of a subset of it. Let $X_n = \{x_0, \ldots, x_n\}$ for $n \in \omega$, $C_n = \mathrm{CN}(X_n)$. Now $y \in \mathrm{CN}(X_n)$, and hence, since A_α is closed and $y \not\in A_\alpha$, we have $\mathrm{CN}(X_n) \not\subset F(X_n)$ and then, by the definition of F, $|\mathrm{CN}(X_n)| > \omega$. This contradicts the assumption that $H_{\omega,\omega} + \Phi \not\subset G$. Now, by the induction hypothesis, $H_{\omega,\omega} + \Phi \not\subseteq G$ implies that for each $G[B_\alpha]$, there exists an ω-well-ordering $<_\alpha$. Define the well-ordering $<$ by $B_\alpha < B_\beta$ for $\alpha < \beta$, and $< = <_\alpha$ on B_α for $\alpha, \beta < \kappa$. Then $N(<, x) = (N(x) \cap A_\alpha) \cup N(<_\alpha, x)$ for $x \in B_\alpha$. Hence $N(<, x)$, a union of two finite sets, is finite for $x \in V$. □

It is proved in Hajnal and Komjáth (1984) that there is a graph of size $2^{\aleph_0}$ with $\chi(G) > \omega$ not containing $H_{\omega,\omega} + 2$. We do not give a proof here.

Next we give the proof of the theorem of Erdős et al. (1972) already mentioned.

Theorem 6.4. *Assume $\chi(G) > \omega$. There is an $r < \omega$ such that $C_s \subset G$ for all $s > r$.*

Proof. First we mention two elementary facts: (i) assume $V = \bigcup_{\beta<\alpha} V_\beta$, $G = \langle V, E \rangle$; then

$$\chi(G) \leqslant \sum_{\beta<\alpha} \chi(G[V_\beta]) , \tag{6.4a}$$

and (ii) assume $G_\beta : \beta < \alpha$ are graphs on the same vertex set $G = \bigcup_{\beta<\alpha} G_\beta$; then

$$\chi(G) \leqslant \prod_{\beta<\alpha} \chi(G_\beta) . \tag{6.4b}$$

Let $G = \langle V, E \rangle$, $\chi(G) > \omega$ be given. We may assume that G is connected. Let $x \in V$, and let A_i be the set of y with distance i from x. By (6.4a), there is an $i \in \mathbb{N}$ with $\chi(G[A_i]) > \omega$. Let $G[A_i] = G'$, $A_i = V'$. Note that for any pair $y \neq z \in V'$ there is a walk of length $2i$ from y to z with inner points outside V'; hence there is a path of length $2j$, $j = j(y, z)$, $1 \leqslant j \leqslant i$, with the same property. Now split G' into the union of G'_j according to what $j(y, z)$ is. By (6.4b), for some j, $\chi(G'_j) > \omega$, and hence, by Theorem 6.3, for every $k \geqslant 2$ there is a C_{2k} in G'_j. Replacing an arbitrary edge $\{y, z\}$ of this C_{2k} with the even path of length $2j$, we get an odd circuit of length $2j + 2k - 1$ for $k \geqslant 2$. □

Note that using the stronger form of Theorem 6.3, we get an edge of G which is contained in all circuits of length $\geq r$ for some $r \in \omega$.

We now turn to the examples mentioned before.

Let $(A, <)$ be an ordered set, $r \geq 2$. The *r-shift-graph on* $(A, <)$ is a graph $\mathrm{Sh}_r(A, <)$ with vertex set $[A]^r$. Two vertices $\{x_0, \ldots, x_{r-1}\}$, $\{y_0, \ldots, y_{r-1}\} \in [A]^r$ are joined in $\mathrm{Sh}_r(A, <)$ if and only if $x_0 < \cdots < x_{r-1}$, $y_0 < \cdots < y_{r-1}$, and either $y_0 = x_1, \ldots, y_{r-2} = x_{r-1}$ or $x_0 = y_1, \ldots, x_{r-2} = y_{r-1}$.

We also define the *2-shift-graph of a graph* $G = \langle V, E\rangle$ *with respect to the ordering* $<$ of V. This will be denoted by $\mathrm{Sh}(G, <)$. The vertex set of this graph is E, while two pairs $\{x, y\}, \{z, w\} \in E$, $x < y$, $z < w$, are joined in $\mathrm{Sh}(G, <)$ if and only if either $y = z$ or $x = w$. Note that $\mathrm{Sh}(G, <)$ is a subgraph of the line graph $L(G)$ of G. The following lemma establishes a connection between the two concepts defined above.

Lemma 6.5. *Let* $r \geq 2$, *and let* $(A, <)$ *be an ordered set. There is an ordering* $<_r$ *of* $[A]^r$ *such that*

$$\mathrm{Sh}_{r+1}(A, <) = \mathrm{Sh}(\mathrm{Sh}_r(A, <), <_r)\,.$$

Proof. Identify $[A]^r$ with the set of $<$ increasing sequences of length r from A and choose $<_r$ as the lexicographic ordering. □

Lemma 6.6. *Assume* $G = \langle V, E\rangle$ *does not contain a* C_{2j+1} *for* $j \leq k$, *for some* $k \geq 0$. *Then* $\mathrm{Sh}(G, <)$ *does not contain a* C_{2i+1} *for* $i \leq k + 1$ *for any ordering* $<$ *of* V.

Proof. Assume $e_0, \ldots, e_{2i}$ is a circuit of $\mathrm{Sh}(G, <)$ without a chord, for some $i \leq k + 1$. It is easy to see that by deleting all e_j for which e_{j-1} and e_{j+1} have the same point in common with e_j, we obtain a circuit C of G. An e_j was deleted if and only if either its upper endpoint is a local minimum of C or its lower endpoint is a local maximum of C in the ordering $<$, and both cannot hold for any e_j. Hence the number of deleted e_j is even. □

Corollary 6.7. *Assume* $r \geq 2$, *and let* $(A, <)$ *be any ordered set. The odd girth of* $\mathrm{Sh}_r(A, <)$ *is at least* $2r + 3$.

Proof. By induction on r. It is clear that $\mathrm{Sh}_2(A, <)$ does not contain a K_3. Assume $r \geq 2$ and that the statement is true for $\mathrm{Sh}_r(A, <)$. Then, by Lemmas 6.5 and 6.6 it is true for $\mathrm{Sh}_{r+1}(A, <)$ as well. □

Theorem 6.8. *Assume* $|A| > \exp_{r-1}(\kappa)$ *for* $r \geq 2$ *and* $\kappa \geq \omega$. *Then* $\chi(\mathrm{Sh}_r(A, <)) > \kappa$ *for any ordering* $<$ *of* A.

Proof. Assume $f : [A]^r \to \kappa$ is a good coloring of $[A]^r$, which is the vertex set of $\mathrm{Sh}_r(A, <)$. By the Erdős–Rado theorem 4.8, there is a subset $\{x_0, \ldots, x_r\} \subseteq A$,

$x_0 < \cdots < x_r$, homogeneous for f. But then $f(\{x_0, \ldots, x_{r-1}\}) = f(\{x_1, \ldots, x_r\})$, and $\{x_0, \ldots, x_{r-1}\}$ is adjacent to $\{x_1, \ldots, x_r\}$ in $\mathrm{Sh}_r(A, <)$, a contradiction. □

As a corollary of Corollary 6.7 and Theorem 6.8, for every $r \geq 2$ and $\kappa \geq \omega$, we obtain a graph with odd girth greater than $2r+1$ and having chromatic number greater than κ. Since the cardinality of this graph is much larger than κ^+, we must give another example as well. I described the shift-graphs in detail, because, I think they are quite basic in both finite and infinite combinatorics. They provide an easy answer to quite a few problems. To illustrate this point I prove a theorem of Erdős and Hajnal mentioned earlier.

Theorem 6.9. *Assume $\kappa \geq \omega$. There is a graph G with $\chi(G) > \kappa$ on $(2^\kappa)^+$ vertices, such that every subgraph G' with 2^κ vertices has chromatic number at most κ.*

Proof. Choose G as $\mathrm{Sh}_2((2^\kappa)^+, <)$. By Theorem 6.8, we have $\chi(G) > \kappa$. To prove the second property of G we need the following lemma. □

Lemma 6.10. *Assume $(A, <)$ is an ordered set, and $|A| \leq 2^\kappa$ for some $\kappa \geq \omega$. Then* $\mathrm{Sh}_2(A, <)$ *has chromatic number at most κ.*

For the proof of this lemma we need another lemma. Let $\mathcal{F}$ be a family of sets. $\mathcal{F}$ is said to be a Sperner family if $A \not\subseteq B$ for any pair $A \neq B \in \mathcal{F}$.

Lemma 6.11. *Assume $\kappa \geq \omega$. Then there is a Sperner family $\mathcal{F}$, $|\mathcal{F}| = 2^\kappa$, consisting of the subsets of κ.*

Proof. It is sufficient to prove this lemma choosing any underlying set of size κ. Hence, let $X = \kappa \times 2$. For any $A \subseteq \kappa$, let $F_A = \{(a, 1) : a \in A\} \cup \{(a, 0) : a \in \kappa \setminus A\}$. It is clear that $\mathcal{F} = \{F_A : A \subseteq \kappa\}$ is a Sperner family of size 2^κ of subsets of X. □

Proof of Lemma 6.10. Let $\{F_a : a \in A\}$ be a one-to-one enumeration of some Sperner family of size $\leq 2^\kappa$ of subsets of κ. For $\{a, b\} \in [A]^2$, $a < b$, define $f(\{a, b\}) = \min_<(F_a \setminus F_b)$. Then f is a coloring of the vertex set of $\mathrm{Sh}_2(A, <)$ with κ colors. We prove that f is a good coloring. Let $a < b < c$, $a, b, c \in A$, be an edge of the shift graph $\mathrm{Sh}_2(A, <)$. Then $f(\{a, b\}) \not\in F_b$ and $f(\{b, c\}) \in F_b$, hence $f(\{a, b\}) \neq f(\{b, c\})$. □

We now define the generalized Specker graphs $\mathrm{Sp}_{r,s}(\kappa)$ for $r \geq 3$, $s < r$, and $\kappa \geq \omega$. The vertex set of $\mathrm{Sp}_{r,s}(\kappa)$ is the set of increasing sequences of length r with values from κ. This will be denoted by $V_{r,s}(\kappa)$. Let $x = (x_0, \ldots, x_{r-1})$ denote the general element of $V_{r,s}(\kappa)$. Let $x, y \in V_r(\kappa)$. $x_0 < y_0$, x and y are joined in $V_{r,s}(\kappa)$ if

$$x_s < y_0 < x_{s+1} < y_1 < \cdots < x_r < y_{r-s}.$$

Theorem 6.12. *Assume $\kappa \geqslant \omega$ is regular, $i \in \mathbb{N}$. Then* $\mathrm{Sp}_{2i^2+i,i}$ *is a graph of cardinality and chromatic number κ with odd girth at least $2i+3$.*

Proof (*Sketch*). As to the odd girth of the graph, for $i=1$, it is very easy to see that $S_{3,1}(\kappa)$ does not contain a triangle. Indeed, if x, y, z are three vertices with $x_0 < y_0 < z_0$ and x is adjacent to y and z', then both y_0 and z_0 are in the interval (x_1, x_2); hence y and z are not joined. This was Specker's original idea.

The general statement is a cumbersom exercise in finite combinatorics and is left to the reader.

The claim for the chromatic number follows from the following facts. We define *full size* subsets of V_r for $r \in \mathbb{N}$. $A \subseteq V_1$ is full size if $|A| = \kappa$. $A \subset V_{r+1}$ is full size if

$$\{x_0 \in \kappa : \{(x_1, \ldots, x_r) \in V_r : (x_0, \ldots, x_r) \in A\} \text{ is full size in } V_r\}$$

has cardinality κ.

Prove first that for every coloring of V_r with fewer than κ colors there is a monochromatic full size subset A of V_r. Finally prove that if A is a full size subset of V_r, then there are $(x_0, \ldots, x_{r-1})$, $(y_0, \ldots, y_{r-1}) \in A$ for any prescribed pattern $x_0 < \cdots < x_{i_0} < y_0 < \cdots < y_{j_0} < \cdots$. □

Appendix

$\{x\}$, $\{x, y\}$, $\{x, y, z\}$, etc., denote sets with elements x; x and y; x, y and z, and so on, respectively. $(x, y) = \{\{x\}, \{x, y\}\}$, $(x, y, z) = ((x, y), z)$, etc., are ordered pairs and ordered triples, respectively. $\{x \in A : \Phi(x)\}$ denotes the set of elements of A having property Φ. Relations and functions are sets consisting of ordered pairs. $\mathrm{Dom}(A)$ and $\mathrm{Ran}(A)$ are the sets of first and second elements of ordered pairs contained in A, denoted the domain and the range of a function, respectively.

The sets A and B are *equipollent* if there is a one-to-one mapping f of A onto B. We write $A \sim B$ in this case.

The pair $(A, <)$ is an *ordered set* if $<$ is an irreflexive and transitive relation on A so that exactly one of $a < b$, $a = b$, $b < a$ always holds. An ordered set $(A, <)$ is *well-ordered* if every non-empty subset of A has a $<$-minimal element. The ordered sets $(A, <)$ and $(A', <')$ are *isomorphic* or *similar* if there is a one-to-one mapping of A onto A' which is strictly monotonic. We write $(A, <) \cong (A', <')$ in this case.

We will need some specific sets called ordinals. A set A is called an *ordinal number* or an *ordinal* for short if A consists of sets, every element of A is a subset of A and ϵ well-orders A. The last statement means more precisely the following. If $\epsilon_A = \{(x, y) : x, y \in A \text{ and } x \in y\}$, then (A, ϵ_A) is a well-ordered set. For example, it is easy to see that $\emptyset$, $\{\emptyset\}$, $\{\emptyset, \{\emptyset\}\}$, $\{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}, \ldots$ are ordinals.

In what follows $\alpha, \beta, \gamma, \ldots, \mu, \nu, \xi, \zeta$ run over ordinals. Define $\alpha < \beta$ by $\alpha \in \beta$. The following are easy consequences of the definitions.

Corollary A.1. *$<$ is a well-ordering of the ordinals, every element of an ordinal is an ordinal, each ordinal is the set of all smaller ordinals, $\alpha < \beta$ is equivalent to $\alpha \subset \beta$, $\alpha \leqslant \beta$ is equivalent to $\alpha \subseteq \beta$.*

For α an ordinal, define $\alpha \dotplus 1 = \alpha \cup \{\alpha\}$. $\alpha \dotplus 1$ is the smallest ordinal greater than α. β is a *successor* ordinal if $\beta = \alpha \dotplus 1$ for some α. A non-successor ordinal different from 0 is a *limit ordinal*. An ordinal β is *finite* if it is not a limit ordinal and the same holds for every element of β. The set of all finite ordinals is denoted by ω. We identify the non-zero elements of ω with the natural numbers

$$\omega = \mathbb{N} \cup \{0\} .$$

The reader should be warned that we only invoke the above convention if it is suitable for the particular purpose at hand.

A set is finite if it is equipollent with a finite ordinal. It can be proved (using the axiom of choice) that a set is finite if and only if it is not equipollent to any proper subset of itself.

The next theorem describes the concept of cardinality.

Theorem A.2. *There is an operation associating with each set A an ordinal $|A|$, called the* cardinality *of A, such that*:

$|A|$ *is the smallest ordinal equipollent with A.*

The possible values of $|A|$ are the *cardinals*.

The cardinalities of finite sets are the *finite cardinals*; the rest are the infinite cardinals. By Corollary A.1, the finite cardinals are the non-negative integers. ω is the smallest infinite cardinal. A set A is called countable if $|A| \leqslant \omega$. $i, j, k, l, m, n, r, s, t$ run over non-negative integers. $\kappa, \lambda, \rho, \sigma, \tau$ run over cardinals.

The next theorem describes the concept of order type.

Theorem A.3. *There is an operation associating to each ordered set $(A, <)$, an object* typ $A(<)$ *called* the order type of A in the ordering $<$ *such that for every pair $(A, <)$, $(A', <')$ of ordered sets $(A, <) \cong (A', <')$ if and only if* typ $A(<) =$ typ $A'(<')$, *and for every well-ordered set $(A, <)$,* typ $A(<)$ *is the unique ordinal α with*

$$(A, <) \cong (\alpha, \epsilon_{\alpha}) .$$

Note that cardinals being ordinals, have an ordering inherited from the ordering of ordinals. The following result shows that this is the same as the traditional ordering.

Theorem A.4. *For any sets A, B $|A| < |B|$ if and only if A is equipollent to a subset of B, but A and B are not equipollent.*

Addition, multiplication and exponentiation of non-negative integers can be extended to infinite cardinals in such a way that:

Fact A.5. *For any pair of disjoint sets A and B, $|A \cup B| = |A| + |B|$. For any pair of sets A and B, $|A| \cdot |B| = |A \times B|$ and $|B_A| = |A|^{|B|}$, where $A \times B$ is the Cartesian product of the sets A and B, and ${}^B A = \{f : f$ is a function mapping B into $A\}$.*

More generally:

Fact A.6. *For any sequence $(A_\alpha : \alpha < \beta)$ of pairwise disjoint sets $|\bigcup_{\alpha<\beta} A_\alpha| = \sum_{\alpha<\beta} |A_\alpha|$, and $|\times_{\alpha<\beta} A_\alpha| = \prod_{\alpha<\beta} |A_\alpha|$ for any sequence of sets $(A_\alpha : \alpha < \beta)$.*

Here $\times_{\alpha<\beta} A_\alpha$ is the Cartesian product of the sets A_α *consisting of all choice functions for the family* $(A_\beta : \beta < \alpha)$, i.e.,

$$f \in \times_{\alpha<\beta} A_\alpha$$

if and only if $\mathrm{Dom}(f) = \beta$ and $f(\alpha) \in A_\alpha$ for all $\alpha < \beta$.

The next result is the fundamental theorem of cardinal arithmetic.

Theorem A.7. *$\kappa^2 = \kappa$ for all $\kappa > \omega$.*

This result has the following corollaries: $\kappa\lambda = \kappa + \lambda = \max\{\kappa, \lambda\}$ provided one of the cardinals κ, λ is infinite, and $\sum_{i\in\omega} \kappa^i = \kappa$ for $\kappa \geqslant \omega$, i.e., the set of all finite sequences formed from a set of cardinality κ is of cardinality κ.

κ^+ denotes the *smallest cardinal* greater than κ. Hence $\kappa^+ = \kappa + 1$ for finite κ. Theorem A.7 easily implies that for $\kappa \geqslant \omega$

$$\kappa^+ = \{\alpha : |\alpha| = \kappa\} ,$$

i.e., κ^+ is the cardinality of the set of ordinals having cardinality κ.

It is easy to see that for a set A of ordinals, $\bigcup A = \bigcup\{\alpha : \alpha \in A\} = \bigcup_{\alpha\in A}$ is the smallest ordinal which is greater than or equal to each element of A. Hence $\bigcup A$ is also denoted by $\sup A$. It is also easy to see that if A consists of cardinals, then $\sup A$ is a cardinal, and as a consequence of Theorem A.7:

$$\sup A = \sum_{\lambda\in A} \lambda \quad \text{provided } \sup A \text{ is infinite} .$$

As to the multiplication, the situation is quite different as proved by J. König.

Theorem A.8. *Let $\lambda_\beta < \kappa_\beta$ for $\beta < \alpha$. Then*

$$\sum_{\beta<\alpha} \lambda_\beta < \prod_{\beta<\alpha} \kappa_\beta .$$

Choosing $\lambda_\beta = 1$, $\kappa_\beta = 2$ we see that this is a generalization of Cantor's classical result $|\alpha| < 2^{|\alpha|}$, and it can be proved by the same diagonal method.

The following concepts are fundamental for investigating cardinal arithmetic. Let $(A, <)$ be an ordered set, $B \subseteq A$. $(A, <)$ is *cofinal* with B if for all $a \in A$ there is a $b \in B$ with $a \leqslant b$. For example, $(\mathbb{R}, <)$ is cofinal with $\mathbb{N}$. The following is a basic theorem due to Hausdorff.

Theorem A.9. *Let $(A, <)$ be an ordered set. There is a subset $B \subset A$ such that B is well-ordered in the ordering inherited from A, A is cofinal with B, and*

$$\operatorname{typ} B(<) \leqslant |A| .$$

On the basis of this theorem one can define $\operatorname{cf}(A, <)$, the *cofinality of an ordered set* $(A, <)$, as the smallest ordinal α such that $(A, <)$ has a cofinal subset of type α. It is a useful fact to know that if B is a cofinal subset of $(A, <)$, then $\operatorname{cf}(A, <) = \operatorname{cf}(B, <)$. Since the cofinality of similar ordered sets is the same, $\operatorname{cf}(\theta)$, *the cofinality of the order type* θ, can be defined. When θ is an infinite cardinal, the cofinality has the following characterization.

Theorem A.10. *For $\kappa \geqslant \omega$, $\operatorname{cf}(\kappa)$ is the smallest cardinal λ such that $\kappa = \sum_{\beta<\lambda} \kappa_\beta$ for a sequence of cardinals $(\kappa_\beta : \beta < \lambda)$ with $\kappa_\beta < \kappa$ for $\beta < \lambda$, i.e., κ is the sum of λ many cardinals all of which are smaller than λ.*

Since $\kappa = \sum_{\beta<\kappa} 1$, we clearly have $\operatorname{cf}(k) \leqslant \kappa$. Let κ be an infinite cardinal. κ is called *regular* if $\operatorname{cf}(\kappa) = \kappa$; otherwise it is *singular*. For example, ω is a regular cardinal, since the sum of finitely many finite numbers is finite. It is an easy theorem that for an arbitrary order type $\theta \neq 0$ either $\operatorname{cf}(\theta) = 1$ or $\operatorname{cf}(\theta)$ is a regular cardinal. A cardinal κ is a *successor* cardinal if $\kappa = \lambda^+$ for some λ; otherwise it is a *limit* cardinal. It follows from Theorem A.7 that every infinite successor cardinal is regular. ω is a regular limit cardinal. A regular limit cardinal is called *inaccessible*. This expression was introduced since, by definition, these cardinals cannot be obtained from smaller cardinals using the successor operation and addition. If, in addition, this holds for the exponentiation too, i.e., if $2^\lambda < \kappa$ for all $\lambda < \kappa$ for an inaccessible cardinal κ, then κ is called *strongly inaccessible*. Clearly, ω satisfies this requirement as well.

It cannot be proved from the axioms that there are inaccessible cardinals different from ω. The assumption that such cardinals exist is an extension of the axiom-system of set theory, and is called a large cardinal axiom. We discussed more about this in section 5.

Induction and recursion can be generalized for ordinals. The following is a formal statement of this fact.

Theorem A.11. (i) *Let $\Phi(\alpha)$ be a property of ordinals. Assume that for each α, if $\Phi(\beta)$ holds for all $\beta < \alpha$, then $\Phi(\alpha)$ is true as well. Then $\Phi(\alpha)$ holds for every α.*

(ii) *Assume $G(A)$ is an operation associating sets to sets. Then there exists an operation $\mathcal{F}(\alpha)$ defined uniquely for all ordinals α such that*

$$\mathcal{F}(\alpha) = G(\mathcal{F} \restriction \alpha) \quad \textit{for all } \alpha .$$

Here $\mathcal{F} \restriction \alpha$ is the restriction of $\mathcal{F}$ to α, i.e., $\mathrm{Dom}(\mathcal{F} \restriction \alpha) = \alpha$ and $\mathcal{F} \restriction \alpha(\beta) = \mathcal{F}(\beta)$ for each $\beta < \alpha$.

Using Theorem A.11, one can define an operation on ordinals listing all infinite cardinals in increasing order. Cantor introduced this operation, and denoted it by $\aleph$ (aleph), the first letter of the Hebrew alphabet. So $\aleph_0, \aleph_1, \ldots, \aleph_\alpha, \ldots$ list all infinite cardinals in increasing order. It is clear from the definitions that $\aleph_\alpha$ is a successor cardinal if and only if α is a successor ordinal, e.g., $\aleph_0, \aleph_1, \ldots, \aleph_n, \ldots$ are regular cardinals and $\aleph_\omega$ is a singular cardinal with $\mathrm{cf}(\aleph_\omega) = \aleph_0 = \omega$. Note that many papers use ω_α as an alternative notation for $\aleph_\alpha$, and we will follow this practice.

One of the few old results on cardinal exponentiation is the following theorem, successively developed by Bernstein, Hausdorff and Tarski.

Theorem A.12. *Assume* $\kappa \geqslant \omega$, $0 < \lambda < \mathrm{cf}(\kappa)$. *Then*

$$\kappa^\lambda = \Big(\sum_{\tau < \kappa,\ \tau \text{ a cardinal}} \tau^\lambda \Big) \kappa .$$

The continuum hypothesis is the assumption that $2^{\aleph_0} = \aleph_1$. This will be denoted by CH. The generalized continuum hypothesis (GCH) is the assumption that $2^\kappa = \kappa^+$ holds for all infinite κ, or equivalently, $2^{\aleph_\alpha} = \aleph_{\alpha+1}$ holds for all α. It is well known that both are consistent with the axioms of set theory and neither of them can be proved from the system, provided the system itself is consistent.

For the time being, we only mention that if the GCH is assumed, we can compute the cardinal power κ^λ for all infinite κ. Using Theorems A.8 and A.12, it is easy to see that

$$\begin{aligned} &\kappa^0 = 1\,, && \text{for } 0 < \lambda < \mathrm{cf}(\kappa)\,, \\ &\kappa^\lambda = \kappa^+ && \text{for } \mathrm{cf}(\kappa) \leqslant \lambda \leqslant \kappa\,, \\ &\kappa^\lambda = \lambda^+ && \text{for } \lambda \geqslant \kappa\,. \end{aligned} \tag{A.13}$$

For a long time it was thought that Theorems A.8 and A.12 contain all information for cardinal exponentiation one can obtain, without additional assumptions. This is not so, and there are quite a few additional non-trivial inequalities. We refer the reader to Silver (1975), Galvin and Hajnal (1975), and Shelah (1982).

We will assume that the reader is familiar with all forms of Zorn's lemma, and the well-ordering theorem. We omit giving a list of them. We only state an important consequence. Let an underlying set X be fixed. A family $\mathcal{F}$ of subsets of X is said to have the *finite intersection property* if $\bigcap \mathcal{F}' \neq \emptyset$ for all finite $\mathcal{F}' \subset \mathcal{F}$. A family U of subsets of X is a *filter* if $U \neq \emptyset$, $\emptyset \in U$, and $A \in U$ and $A \subseteq B$ imply $B \in U$, and U is closed with respect to finite intersections. A filter U is an *ultrafilter* if for each $A \subset X$ either A or $X \backslash A$ belongs to U.

Theorem A.14. *If $\mathcal{F}$ is a family of subsets of X having the finite intersection property, then $\mathcal{F}$ extends to an ultrafilter in X.*

It is easy to see that any extension of $\mathcal{F}$, which is maximal for having the finite intersection property, satisfies the requirements of Theorem A.14.

An ultrafilter is *free* or *non-principal*, if $\{x\} \not\in U$ for every $x \in X$. Since the cofinite sets of an infinite set X form a filter, it follows from Theorem A.14 that every infinite set carries a free ultrafilter.

Finally, we will need an important tool of elementary set theory which is probably less well known than the ones listed above.

Definition A.15. Let $\kappa > \omega$ be a regular cardinal. $B \subseteq \kappa$ is a *club-set* (in κ) if κ is cofinal with B, and B is closed in the order topology of κ, i.e., for all $B' \subset B$, $B' \neq 0$ if $B' \subset \alpha$ for some $\alpha < \kappa$, $\sup B' \in B$. The term "club" comes from closed unbounded.

For each $\alpha < \kappa$, $\kappa \setminus \alpha$ is a club-set. However, there are club-sets with a complement of size κ. For example, $\{\alpha < \kappa : \alpha$ is a limit ordinal$\}$ is such a set. Still, club-sets are very large as is shown by:

Lemma A.16. *If $\mathcal{F}$ is a family of club-sets in κ and $|\mathcal{F}| < \kappa$, then $\bigcap \mathcal{F}$ is a club-set in κ for $\kappa = \mathrm{cf}(\kappa) > \omega$.*

By this lemma, the sets of κ which contain a club-set form a filter, and the intersection of fewer than κ sets in this filter belongs to the filter. We call such filters *κ-complete*. We think of sets in the filter as large sets and their complements as small sets.

$A \subset \kappa$ is *stationary* in κ if $A \cap B \neq \emptyset$ for all club-sets B in κ. By Lemma A.16, all club-sets are stationary. Stationary subsets of κ are those, which are not small, but also not necessarily large. We need a characterization of these sets.

Let $A \subseteq \kappa$ and f be a function such that $A \subseteq \mathrm{Dom}(f)$. Then f is *regressive* or *pressing down* on A if $f(\alpha) < \alpha$ for $\alpha \in A$ and $\alpha \neq 0$.

Lemma A.17. *$A \subseteq \kappa$ is stationary in κ if and only if for all regressive functions f on A, there is a $\beta < \kappa$ with $|f^{-1}(\{\beta\}) \cap A| = \kappa$, i.e., every regressive function on A takes one of its values on A κ many times.*

The following is the main tool for handling these concepts (see Fodor 1956).

Theorem A.18. *Let $\kappa > \omega$ be regular. $A \subset \kappa$ is stationary in κ if and only if for every regressive function f on A, the set $f^{-1}(\{\beta\}) \cap A$ is stationary in κ for some $\beta < \kappa$, i.e., every regressive function on A takes one of its values "stationary many times".*

We will use the following elementary fact.

Fact A.19. *Assume $\lambda < \kappa$ are regular cardinals. Then the set $A_{\kappa,\lambda} = \{\alpha < \kappa$: $\mathrm{cf}(\alpha) = \lambda\}$ is a stationary subset of κ.*

As we have already mentioned, stationary sets are not necessarily large. There are stationary sets whose complement is stationary. For example, by Fact A.19, for $\kappa = \omega_2$, $A_{\omega_2,\omega}$ and A_{ω_2,ω_1} are both stationary.

It is somewhat harder to find two disjoint stationary subsets of ω_1, since in this case the above simple minded argument does not work. In general, the following result of Solovay (1971) holds.

Theorem A.20. *Assume $\kappa > \omega$ is regular, and $A \subseteq \kappa$ is stationary in κ. Then A is the disjoint union of κ-many sets, each stationary in κ.*

We are not going to use this result, which is not easy to prove. We only stated it to show that Lemma A.16 can not be generalized to stationary sets. I want to emphasize the importance of the concepts defined in A.15 once more. Countably complete filters having a natural definition played an important role in the development of various branches of mathematics. Subsets of Lebesgue measure 1 of [0, 1], and comeager sets are the prime examples. Club-sets and stationary sets play a role in set theory similar to the role of Lebesgue measure in the theory of real functions.

References

Banach, S., and K. Kuratowski

[1929] Sur une généralisation du probleme de la mesure, *Fund. Math.* **14**, 127–131.

Baumgartner, J.E.

[1984] Generic graph constructions, *J. Symbolic Logic* **49**, 234–240.

Bollobás, B.

[1978] *Extremal Graph Theory* (Academic Press, New York).

de Bruijn, N.G., and P. Erdős

[1951] A color problem for infinite graphs and a problem in the theory of relations, *Akademia Ámsterdam* **13**, 371–373.

Erdős, P.

[1959] Graph theory and probability, *Canad. Math. J.* **11**, 34–38.

Erdős, P., and A. Hajnal

[1958] On the structure of set mappings, *Acta Math. Acad. Sci. Hungar.* **9**, 111–131.

[1966] On chromatic number of graphs and set systems, *Acta Math. Acad. Sci. Hungar.* **17**, 61–99.

[1968] On chromatic numbers of infinite graphs, in: *Theory of Graphs, Proc. Colloq., Tihany*, ed. V.T. Sós (Akadémiai Kiadó, Budapest) pp. 83–98.

Erdős, P., and R. Rado

[1956] A partition calculus in set theory, *Bull. Amer. Math. Soc.* **62**, 427–489.

Erdős, P., and A. Tarski

[1943] On families of mutually exclusive sets, *Ann. of Math.* **44**, 315–329.

Erdős, P., A. Hajnal and R. Rado

[1965] Partition relations for cardinal numbers, *Acta Math. Acad. Sci. Hungar.* **16**, 93–196.

Erdős, P., A. Hajnal and S. Shelah

[1972] On some general properties of chromatic numbers, in: *Topics in Topology, Keszthely*, ed. A. Császár, *Colloq. Math. Soc. János Bolyai* **8**, 243–255.

Erdős, P., A. Hajnal, A. Máté and R. Rado

[1984] *Combinatorial Set Theory: Partition Relations for Cardinals, Studies in Logic and the Foundations of Mathematics* (Akadémiai Kiadó/North-Holland, Budapest/Amsterdam).

Fodor, G.

[1956] Eine Bemerkung zur Theorie der regressiven Funktionen, *Acta Sci. Math.* **17**, 139–142.

Foreman, M.D., and R. Laver

[1988] Some downward transfer properties for $\aleph_2$, *Adv. in Math.* **67**, 230–238.

Foreman, M.D., M. Magidor and S. Shelah

[1988] Martins maximum, saturated ideals and non-regular ultrafilter, Part I, *Ann. of Math.* **127**, 1–47.

Galvin, F., and A. Hajnal

[1975] Inequalities for cardinal powers, *Ann. of Math.* **101**, 491–498.

Graham, R.L., B.L. Rothschild and J.H. Spencer

[1980] *Ramsey Theory* (Wiley, New York).

Hajnal, A., and P. Komjáth

[1984] What must and what need not be contained in a graph of uncountable chromatic number, *Combinatorica* **4**, 47–52.

Hanf, W.P.

[1964] On a problem of Erdős and Tarski, *Fund. Math.* **53**, 325–334.

Jech, T.

[1978] *Set Theory* (Academic Press, New York).

Kiesler, H.J., and A. Tarski

[1964] From accessible to inaccessible cardinals, *Fund. Math.* **53**, 225–308.

König, D.

[1927] Über eine Schlussweise aus dem Endlichen ins Unendliche, *Acta Sci. Math.* **3**, 121–130.

Kunen, K.

[1977] Combinatorics, in: *Handbook of Mathematical Logic, Studies in Logic and the Foundations of Mathematics*, Vol. 90, ed. I. Barwise (North-Holland, Amsterdam).

Mahlo, P.

[1911] Über lineare transfinite Mengen, *Ber. Verh. Königl. Sächs. Ges. Wissensch. Leipzig, Math.-Phys. Kl.* **63**, 187–225.

Ramsey, F.P.

[1930] On a problem of formal logic, *Proc. London Math. Soc. (2)* **30**, 264–286.

Rowbottom, F.

[1971] Some strong axioms of infinity incompatible with the axiom of constructibility, *Ann. Math. Logic* **13**, 1–44.

Scott, D.S.

[1961] Measurable cardinals and constructible sets, *Bull. Acad. Pol. Sci.* **9**, 521–524.

Shelah, S.

[1982] Proper forcing, *Lecture Notes in Mathematics*, Vol. 940 (Springer, Berlin).

[1990] Incompactness for chromatic numbers of graphs, in: *A Tribute to Paul Erdős*, eds. A. Baker, B. Bollobás and A. Hajnal (Cambridge University Press, Cambridge).

Silver, J.H.

[1966] *Some applications of model theory in set theory*, Doctoral dissertation (University of California, Berkeley, CA). Published: 1971, *Ann. Math. Logic* **3**, 45–110.

[1975] On the singular cardinal problem, in: *Proc. Int. Congr. of Mathematicians, Vancouver, 1974*, Vol. 1, pp. 256–268.

Solovay, R.

[1969] On the cardinality of Σ^1_2 sets of reals, in: *Foundation of Mathematics* (Springer, Berlin) pp. 58–73.

[1971] Real-valued measurable cardinals, in: *Axiomatic Set Theory, Proc. Symp. Pure Math.* **XIII./1**, 397–428.

Ulam, S.

[1930] Zur Masstheorie in der allgemeinen Mengenlehre, *Fund. Math.* **16**, 140–150.

CHAPTER 43

Combinatorial Games

Richard K. GUY

Department of Mathematics and Statistics, University of Calgary, Calgary, Alta. T2N 1N4, Canada

Contents

HANDBOOK OF COMBINATORICS
Edited by R. Graham, M. Grötschel and L. Lovász

1. Introduction

We only skim the surface of this vast subject. For more breadth, depth and detail, consult both of the following books: *Winning Ways for your Mathematical Plays* by Berlekamp et al. (1982) and *On Numbers and Games* by Conway (1976) which we will frequently refer to as WW and ONAG, respectively.

Two other surveys are by Fraenkel (1980), who considers the complexity of games, and Guy (1983), who explores the connexions between games and graphs.

Fraenkel contrasts Nim with Go, the former having a very simple winning strategy and the latter very complicated. Nim has no cycles in its game graph, no interaction between tokens, and is impartial; Go has cycles and interaction, and is partizan. The spectrum between the two games spans the complexity gap between polynomial, P-space-complete and Exptime-complete games. In existential problems such as the travelling salesman problem, high complexity is a liability, but in games and cryptanalysis, it can be an asset.

Fraenkel also maintains a valuable bibliography of the subject, copies of which may be obtained from him at the Weizmann Institute, Rehovot, Israel.

Guy surveys the connexions between combinatorial game theory and graph theory: graphs of games; games on graphs (Hackenbush, von Neumann's game, Rims, Rails, Lucasta, Sprouts); the ways graphs can be used to elucidate puzzles (Tantalizer, Rubik's Cube, Fifteen Puzzle, magic squares); and the occurrence of Euler's formula in Berlekamp's analysis of Dots-and-Boxes (WW, pp. 507–550).

2. What is a game? (WW, pp. 16–17)

Our games, unlike those that have found application in economics, management, and military strategy, are *completely determined*. There is *complete information*: the players know exactly what is going on; there is no bluffing. There are no *chance moves*: no dealing of cards; no rolling of dice. We have only two players, Left and Right: there can be no question of *coalitions*.

For each given *position* in a game, the rules define two sets of *options*, available respectively to Left and to Right, leading to other positions. Left and Right alternately choose an option, i.e., play to a position, specified by the rules.

3. Game graphs and trees

A game may be visualized as a digraph: the nodes are the positions and the arcs are the options. The arcs may be thought of as colored, say

bLue, Red, or grEen,

according as the option is available

to Left only, to Right only, or to Either player.

Alternatively, we may distinguish between different plays of the game, i.e., different dipaths in the digraph, by duplicating the nodes as necessary and representing the game by a rooted tree. The root is the starting position and the arcs are directed away from the root.

Figures 1 and 2 show the game graph and the game tree for the position {3, 2} in a game of Nim: two heaps, one with three beans, the other with two. Nim is an example of an *impartial* game, in every position of which the same set of options is available to either player: think of the arcs in figs. 1 and 2 as being colored green. Nim is played with a number of heaps of beans. The typical option, for

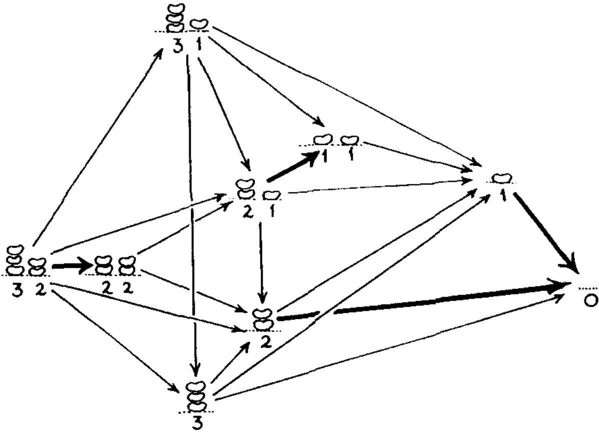


Figure 1. The game graph for the Nim position {3, 2}. From the left-most position the next player can win by adopting the strategy indicated by the heavy arrows.

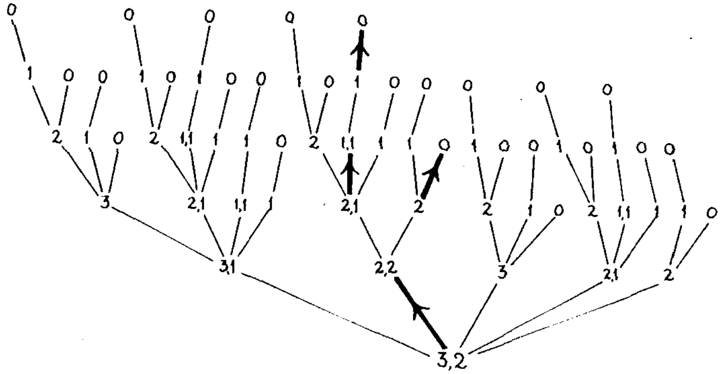


Figure 2. The game tree for the same Nim position. The Root is {3, 2}, and the arcs are directed upwards.

either player, is to choose a heap and remove from it as many beans as you wish: the whole heap maybe, but at least one bean.

Notice the difference between the *complete analysis* of a game, and a *winning strategy*. Figure 2 is a complete analysis: for a winning strategy it suffices to describe the four black arrows.

You may mentally identify three seemingly different aspects of the same idea.

(a) A *game*, i.e., the whole digraph or tree representing the game. For example, *the* Game of Chess, as opposed to *a* game of chess, which we refer to as a *play* of the game: compare *le jeu d'échecs*, *une partie d'échecs*.

(b) A *position* in a game; a particular node of the digraph, perhaps the root of the tree. For example, the standard opening position in Chess, ready for a play of the game.

(c) The *ordered pair* of sets of options available to the two players from a given position, e.g., {a3,a4, . . . , h3,h4,Sa3,Sc3,Sf3,Sh3 | a6,a5, . . . , h6,h5,Sa6,Sc6, Sf6,Sh6}.

A position, such as the rightmost in fig. 1, or any zero in fig. 2, from which neither player has any option, is a *terminal position*, at which the game ends. The *outcome* is then specified by the rules. It may be a win for Left, or a win for Right, possibly accompanied by some score or payoff. The rules may not specify a winner, so that the game may end in a *tie*. For almost all of this chapter, however, we will adopt the *normal play convention* that the winner is the player who has just made the last move: equivalently, *last player winning*; if you cannot move, you lose. Near the end we will say something about the *misère play convention*, which accords the win to a player unable to move: *last player losing*. Analysis is far more difficult in this case.

To ensure that we *have* a last player, our games must end. We assume that they satisfy the *ending condition*: that there is no infinite sequence of options. Notice that this condition prohibits *all* infinite sequences, not merely those in which Left and Right make alternate moves. In order to give *values* to our games, we need to consider the possibility of several consecutive moves by the same player. This can occur in the play of the *sum* of two or more games, as we shall see.

A game that does not satisfy the ending condition is called a *loopy game*. Its digraph will contain a directed circuit or an infinite directed path. The outcome may be a *draw*: note that we distinguish between a *tied* game and one *drawn out* by infinite play. Chess exhibits both kinds of outcome: stalemate is a tie, but perpetual check, repetition of moves, or insufficient mating material are equivalent to draws.

4. The formal definition of a game

This is deceptively simple, each game is an ordered pair of sets of games:

$$G = \{\{G^{L_1}, G^{L_2}, \dots\} \mid \{G^{R_1}, G^{R_2}, \dots\}\}$$

To avoid proliferation of braces, we write this more compactly as

$$G = \{G^L \mid G^R\}$$

where we must remember that G^L and G^R are *sets* of Left and Right options, which may, for example, be infinite, or empty. Indeed the definition is inductive, and the empty set is the basis for the induction, which starts with the *Endgame*

$$\{\emptyset \mid \emptyset\} = \{\ \mid\ \}$$

in which neither player has an option, and which we will denote by 0 (zero).

Here, and from now on, we use familiar symbols, with the strong implication that we can manipulate games in the same way that we manipulate numbers in ordinary arithmetic. Some games behave like numbers and we call them numbers, but to justify the manipulations takes more space than we have here, so turn to (ONAG, pp. 71–96) if you would like more detail and further examples.

It is helpful to attach ordinal numbers, or *birthdays*, to games, and to introduce the idea of *simplicity* (WW, pp. 23–27). When a move is made in a game, it becomes *simpler* in the sense that we arrive at a position with an earlier birthday. All definitions and proofs are inductive in that they are assumed to have been made for all simpler games. The basis is the simplest game of all, the Endgame, born on day zero.

On day one we have two sets, the empty set and the set $\{0\}$ consisting of the Endgame, so that we can visualize 2^2 games. Their game trees (in which Left moves slope up to the left and Right moves slope up to the right) together with their names are shown in fig. 3.

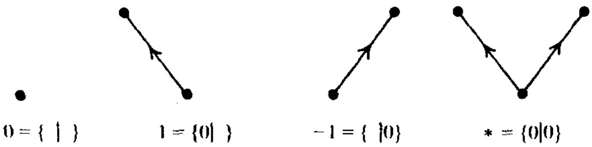


Figure 3. The four simplest games, born on days zero and one.

We quote from (ONAG, p. 72):

> "The simplest game of all is the *Endgame*, 0. I courteously offer you the first move in this game, and call upon you to make it. You lose, of course, because 0 is defined as the game in which it is never legal to make a move.
>
> In the game $1 = \{0 \mid\ \}$, there is a legal move for Left, which ends the game, but at no time is there any legal move for Right. If I play Left, and you Right, and you have first move again (only fair, as you lost the previous game) you will lose again, being unable to move even from the initial position. To demonstrate my skill, I shall now start from the same position, make my legal move to 0, and call upon you to make yours.
>
> Of course you are now beginning to suspect that Left always wins, so for our next game, −1, you may play as Left and I as Right! For the last of our examples, the new

> game $* = \{0 | 0\}$, you may play whichever role you wish, provided that for this privilege you allow me to play first."

In summary:

The Endgame is the prototype of games in which the next player loses, since no option is available: a *second-player win*.

The game 1 is a *Left win*, no matter who starts: if Louise starts, she goes to $\{\ |\ \} = 0$ and Richard has no option and loses; if Richard starts, he has no option and loses even more quickly.

The game -1 is a *Right win*, no matter who starts.

The game $\{0 | 0\} = *$("Star") is the simplest game which is not a number (WW, p. 40). It is a *first-player win*.

5. The four outcome classes

If we adopt the normal play convention, every game belongs to just one of four outcome classes (ONAG, Theorem 50) which are exemplified by the four games we have just seen. The terminology and notation are displayed in fig. 4.

If, in a game G		Right starts	
		& L has a winning strategy	& R has a winning strategy
Left starts	& R has a winning strategy	ZERO $G = 0$ 2nd wins	NEGATIVE $G < 0$ R wins
	& L has a winning strategy	POSITIVE $G > 0$ L wins	FUZZY $G \| 0$ 1st wins

Figure 4. The four outcome classes.

It is convenient to combine these outcome classes and symbols in pairs.

If	Left	Right	Left	Right	has a winning strategy
provided	Right	Left	Left	Right	starts, then we
write	$G \geqslant 0$	$G \leqslant 0$	$G \mathrel{\|\!\rhd} 0$	$G \mathrel{\lhd\!\|} 0$	corresponding to
the	1st col	1st row	2nd row	2nd col	of fig. 4.

6. The negative of a game

A device to breathe new life into an otherwise one-sided contest is to allow a novice opponent, when he feels he is losing, to turn the board around, to reverse the roles of the two players, to handicap his more skilled adversary, by asking her to defend what appears to him to be an inferior position. This replaces the game by its negative. Formally, the *negative* of G,

$$-G = \{G^R \mid -G^L\}$$

is defined inductively (WW, p. 35). Remember that $-G^R$, for example, is short for the set $\{-G^{R_1}, -G^{R_2}, \ldots\}$, whose members are simpler games than $-G$, and have been defined earlier.

7. Sums of games

There are many ways of playing two or more games simultaneously, but often the most natural is what we call the *sum*, or *disjunctive compound* (ONAG, p. 75, WW, p. 33). Nim, for example, is the sum of a number of games of one-heap Nim. In the sum of two or more component games, the player whose turn it is to move selects one component and makes a legal move in it:

$$G + H = \{G^L + H, G + H^L \mid G^R + H, G + H^R\}$$

Once again this is an inductive definition: $G^R + H$, for example, represents the set of options $\{G^{R_1} + H, G^{R_2} + H, \ldots\}$ each of which is a simpler game than $G + H$, so that addition there is already defined.

It is not hard to see that sums are commutative and associative, that $G + 0 = G$, and (ONAG, Theorem 51) that $G + (-G) = 0$. In that last sentence we have used zero in two quite different senses. In $G + 0 = G$ we intended 0 to mean the Endgame, $\{\ \mid\ \}$. In $G + (-G) = 0$ we intended "$= 0$" to mean "is a zero game", i.e., "belongs to the (very large!) equivalence class of games for which the second player has a winning strategy". Check that $1 + (-1) = 0$ and $* + * = 0$, so that we can speak of the games $1 + (-1)$ and $* + *$ as having the same *value*, 0, as that of the Endgame, even though their *forms* are different.

More generally, we will say that two games are *equivalent*, and have the same *value*, and write $G = H$, if the game $G + (-H)$ is a second-player win. With the above definitions of sum, negative and zero, games form a commutative group. Moreover, games are partially ordered, and we write $G > H$ just if $G - H > 0$, i.e., if Left can win the sum $G + (-H)$, no matter who starts. Our notation is justified by theorems such as the following, proved in (ONAG, p. 76). If $G \geqslant 0$ and $H \geqslant 0$, then $G + H \geqslant 0$. If H is a zero game (second player wins), then $G + H$ has the same outcome as G. If $H - K$ is a zero game, then $G + H$ and $G + K$ have the same outcome.

8. The games born on day two

As day two dawns, we have four games to play with, and so $2^4 = 16$ sets of games. So there are 16 choices for Left's options and 16 for Right's giving a potential of 256 games on day two. However, things are not *quite* that complicated, in that, for Left say, some options are clearly preferable to others. The four games born on day one can be arranged in the lattice (in the poset sense of chapter 8, rather than the geometrical sense of chapter 19) of fig. 5, in which Left's preferences are higher, and Right's are lower.

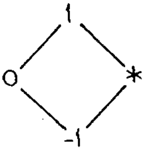


Figure 5. The lattice of games born on day one.

The only set of options for which there is any doubt in either player's mind about the best move, is the incomparable pair $\{0, *\}$. So, for a player's options we need consider only six possibilities: the empty set, the four singletons, and this incomparable pair. Among the resulting 6^2 possibilities for games born on day two, just 22 are inequivalent and 18 are new. They are shown in fig. 6, the four quarters of which should be compared with those of fig. 4. These contain the zero game; six negative games; six positive; and nine fuzzy ones. The six sets of Right

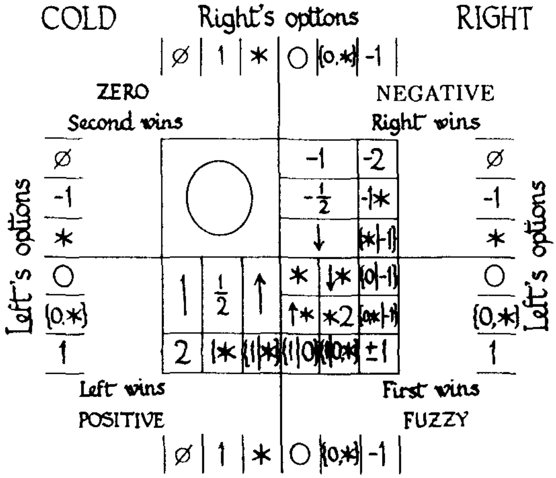


Figure 6. The 22 games born on day two.

options run from left to right in increasing order of desirability from Right's point of view; Left's correspondingly downwards.

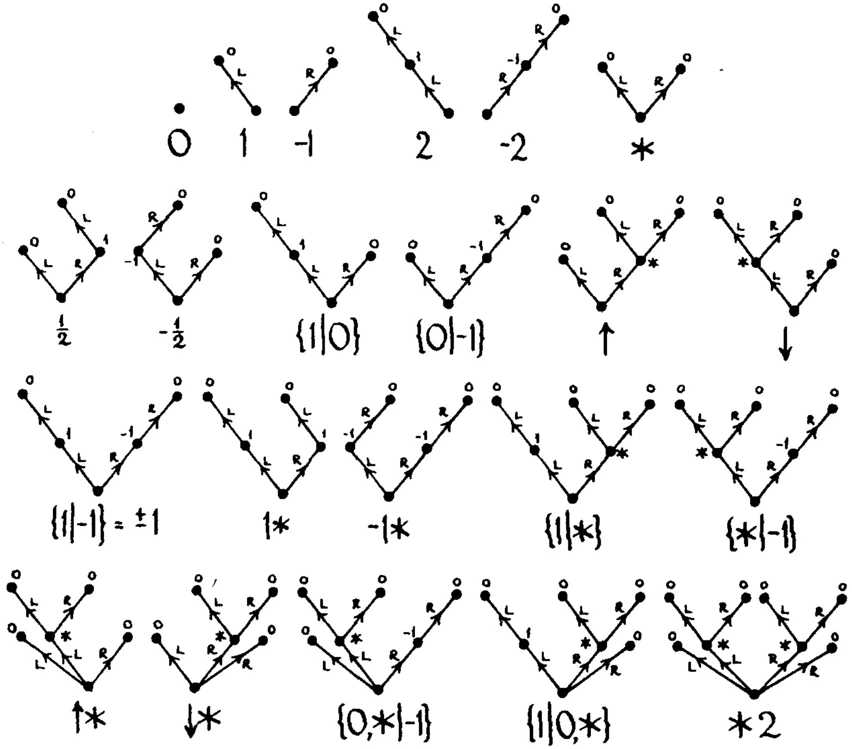


Figure 7. The game trees for the 22 games born on day two.

Figure 7 shows the game trees for these 22 day-two games: the lowest node in each case is the root, arcs sloping up to the left are blue, those sloping up to the right are red. The trees for Star and 16 of the day-two games have been condensed into digraphs in fig. 8; arcs labelled E are green and represent options available to both Left and Right.

The 22 games are exhibited as a lattice in fig. 9. If two games are connected by an arc, or, transitively by a path of arcs, then the higher game has a greater value than the lower, as in fig. 5.

Examples and exercises. $1+1=2$, $\frac{1}{2}+\frac{1}{2}=1$, $*+*=0$, $\uparrow=\{0\,|\,*\}=\{0,*\,|\,*\}>0$ ("Up is positive"), $\uparrow*=\{0,*\,|\,0\}=\uparrow+*$ ("Upstar"), $0\,\|\,*2=\{0,*\,|\,0,*\}$ ("0 is incomparable with Star-two"), $1*=\{1\,|\,1\}=1+*$, $\{1\,|\,0\}>*$, $\downarrow*=\{0\,|\,0,*\}\,\|\,0$ ("Downstar is incomparable with zero"), $\{1\,|\,0\}>\uparrow*$, $\{1\,|\,0\}>*2$, $\{1\,|\,0\}>\downarrow=\{*\,|\,0\}$, $\{1\,|\,0\}>\downarrow*$, $\{1\,|\,0\}>\pm1=\{1\,|\,-1\}$, $\{1\,|\,0\}\,\|\,0$, $\{1\,|\,0\}\,\|\uparrow$, $\{1\,|\,*\}>0$, $\{1\,|\,*\}\mathrel{|\!>}\uparrow$, $\{1\,|\,*\}\,\|\uparrow*$, $\{1\,|\,0,*\}\mathrel{|\!>}\downarrow$, $\{1\,|\,0,*\}\,\|\uparrow$.

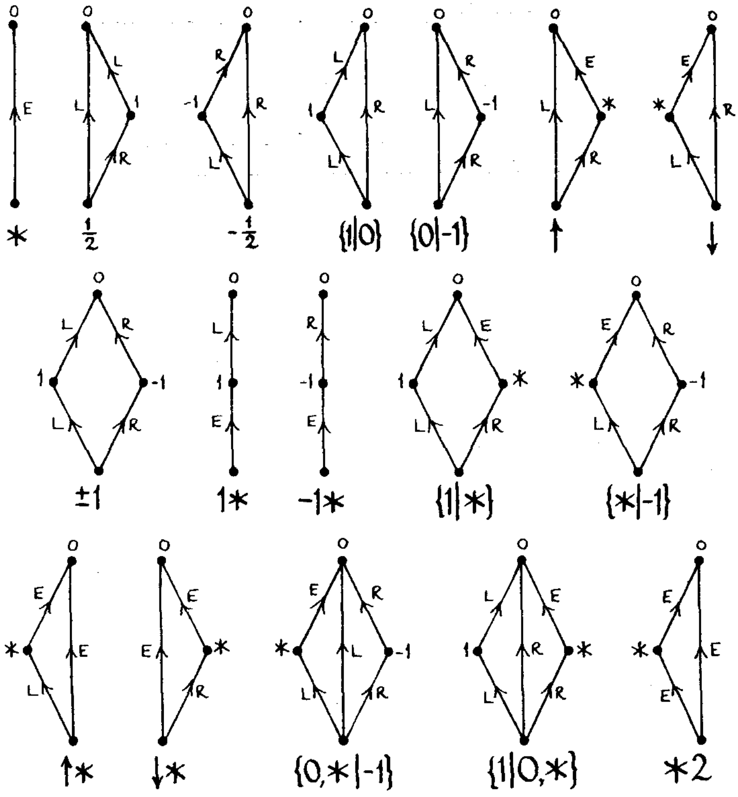


Figure 8. Digraphs for Star and 16 day-two games.

Two important ways of classifying games are as *hot* or *cold* and as *partizan* or *impartial*. We will shortly make a brief attempt to distinguish hot from cold. Impartial games are those in which the two sets of Left and Right options are the same; in partizan games the two sets are different, in general. As illustrative examples, we describe two partizan games: Domineering is a hot game; Blue–Red Hackenbush is a cold one.

9. Domineering

This is also called Crosscram (ONAG, pp. 74–75 and 120–121, WW, pp. 117–120 and 137–140). Left and Right alternately place dominoes so that they exactly cover two squares of a checker-board. Left orients her dominoes North–South and Right puts his East–West. Dominoes must not overlap each other or the edge

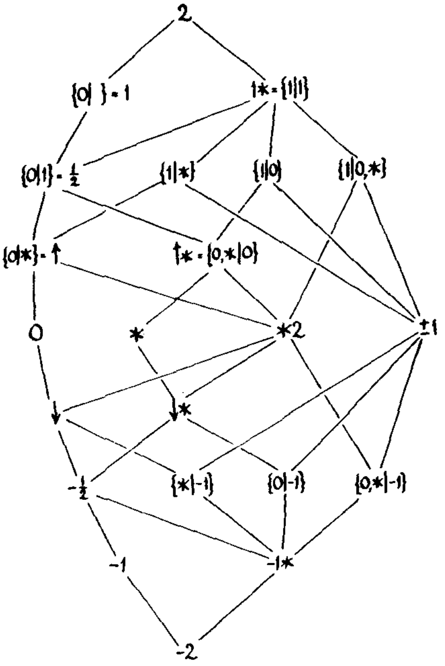


Figure 9. The lattice of day-two games.

of the board. A player loses who can find no appropriately oriented space for a domino. After a while the available space may separate into disconnected regions, and the game becomes the sum of smaller games. Many of the games we have already seen are realized by small Domineering "boards". Check the values in fig. 10.

10. Hot and cold games

Domineering is an example of a *hot game*. These are the interesting games in which there is an advantage in having the move: the first player wins. If $G = \{G^L \mid G^R\}$, then the various differences $G^L - G$ and $G - G^R$ are the (Left and Right) *incentives* of G. These are always negative if G is a number. Numbers are *cold games* and the *Number Avoidance Theorem* tells you:

Never move in a
Number, unless there is
Nothing else to do.

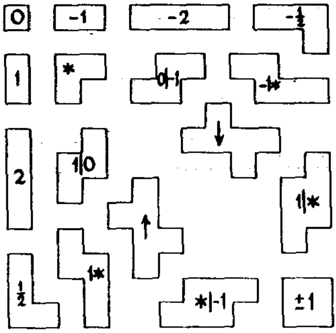


Figure 10. Small Domineering boards realize all 4 day-one games (top left corner) and 13 day-two games.

An earnest of the theory of hot games can be found in the work of Milnor (1953) and Hanner (1959). For recent developments, see (WW, pp. 141–182) and Berlekamp (1988), who is currently generalizing the theory of "overheating" and making inroads into the difficult theory of the game of Go.

A good example of a cold game is given in the next section.

11. Blue–Red Hackenbush

This is perhaps best played on a blackboard, using an eraser. Start with a *picture*, for example fig. 11, which is a graph, some of whose nodes are on the ground (the dotted line), and whose edges are either blue or red (ONAG, pp. 86–91, WW, pp. 3–8). A Left or Right move is to delete a blue or red edge, respectively, together with any edges of either color which are no longer connected to the ground.

If Right deletes the dog's neck, for example, the head disappears as well. If Left deletes the body, no other edges disappear, but the picture breaks into the sum of two separate pictures. The aim, as usual, is to be the last player, the person whose move leaves no edges of the opponent's color.

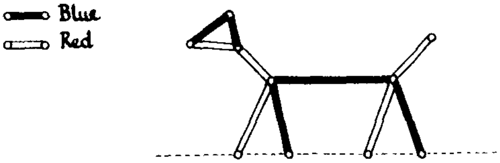


Figure 11. A Blue–Red Hackenbush picture.

12. When is a game a number?

Although the values of Blue–Red Hackenbush pictures may be hard to calculate (in the technical sense; WW, pp. 210–212); they are all numbers. A game is a *number* (ONAG, p. 81) exactly if all its options are numbers and no Left option is $\geq$ any Right option. The game $\pm 1 = \{1 \mid -1\}$ is not a number, since $1 \geq -1$; Star is not a number, because $0 \geq 0$; and $\uparrow = \{0 \mid *\}$ is not a number, because $*$ is not. Examples of numbers are:

0	born on day 0 ,
1 and -1	born on day 1 ,
$\frac{1}{2}, -\frac{1}{2}, 2$ and -2	born on day 2 ,
$\frac{1}{4}, -\frac{1}{4}, \frac{3}{4}, -\frac{3}{4}, 1\frac{1}{2}, -1\frac{1}{2}, 3$ and -3	born on day 3 .

On day ω *all* the remaining real numbers are born, as well as the first infinite ordinals, $\omega = \{0, 1, 2, \ldots \mid \}$ and $-\omega = \{ \mid 0, -1, -2, \ldots\}$ and infinitesimals such as $1/\omega = \{0 \mid 1, \frac{1}{2}, \frac{1}{4}, \ldots\}$.

Values of games may be thought of as "number of moves advantage to Left". For example, -2 is two moves advantage to Right. The first four Blue–Red Hackenbush values in fig. 12 are clear. Deletion of the blue edge in the fifth reduces the picture to 0, while deletion of the red edge leaves 1, so the fifth value is $\{0 \mid 1\} = \frac{1}{2}$.

Check that if you play a game comprising two copies of this and a single separate red edge, fig. 13, then the second player wins. (Although if Left starts, Right *can* make a bad reply!) Check the remaining values in fig. 12.

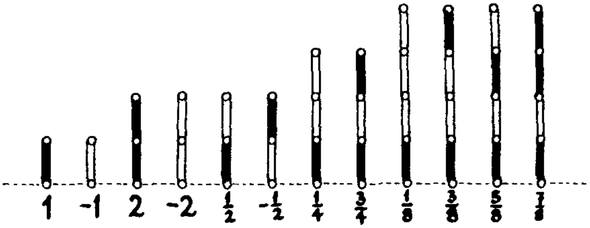


Figure 12. Values of some Blue–Red Hackenbush strings.

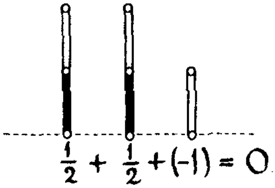


Figure 13.

Not only is the value of every Blue–Red Hackenbush picture a number, but every number can be represented by a Hackenbush string! To see this, work backwards from *Berlekamp's Rule* for evaluating Hackenbush strings (Berlekamp 1974):

> "The sign is determined by the color of the edge touching the ground (+ for blue, − for red). Move up the string until there is a change in edge color. This first *pair* of differently colored consecutive edges represents the *binary point*. The number of edges *below* this pair gives the integer part of the number. Above the pair, label each edge with a binary digit, 1 or 0, according as its color agrees or disagrees with the ground color, and adjoin an extra digit 1 if the string is finite."

The rule is illustrated in fig. 14; also use it to check the values in fig. 12. Infinitesimals and infinite ordinals can also be represented by (infinite) Hackenbush strings (WW, pp. 309–313).

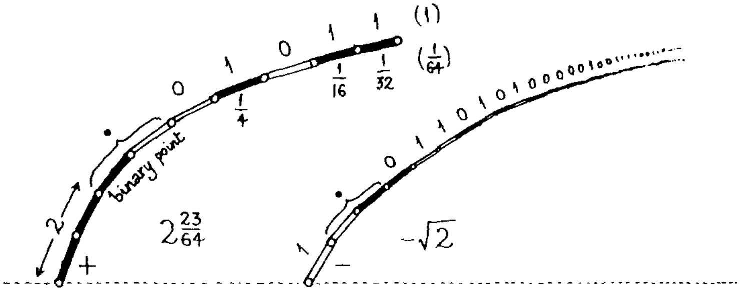


Figure 14. Berlekamp's Rule for Hackenbush strings.

13. Simplifying games

We may be able to simplify a game, in either of two ways (ONAG, pp. 109–112, WW, pp. 62–64):

by deleting dominated options or by replacing reversible options.

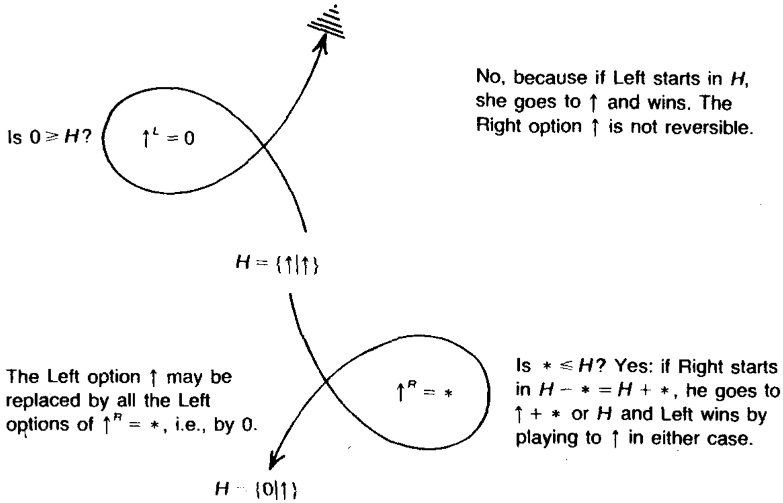
We used the former implicitly when we made our catalog of day-two games. If in a game $G = \{A, B, C, \ldots \mid Z, Y, X, \ldots\}$ we have $B \geqslant A$ (respectively $Y \leqslant Z$), then B *dominates* A (Y dominates Z) and A (Z) may be deleted, provided $B(Y)$ is retained.

Replacing reversible options is more subtle. A right option X is *reversible* if Left has a reply X^L which is at least as good for her as the original game, that is, if $X^L \geqslant G$. Then X may be replaced by the list of *all* the Right options, $X^{LR} = \{X^{LR_1}, X^{LR_2}, \ldots\}$, of X^L. Similarly, the Left option C is reversible if Right has a reply C^R at least as good for him as G is, that is if $C^R \leqslant G$, and C may be replaced by the list of all the Left options C^{RL} of C^R.

For example, in the game

$$G=\{0,*\mid *\}$$

which is more precise description of the shape labelled "$\uparrow$" in fig. 10, neither Left option dominates the other, since $*\,\|\,0$, but the Left option $*$ is reversible, because $*^R=0\leq G$. (To see that $G\geq 0$, note that if Right starts, his only option is $*=\{0\mid 0\}$ and Left plays to 0 and wins.) So the Left option $*$ may be replaced by all the Left options of $*^R=0=\{\ \mid\ \}$. That is, it may be replaced by all the members of the empty set, i.e., it may be deleted, and G simplifies to $\{0\mid *\}=\uparrow$.



No, because if Left starts in H, she goes to $\uparrow$ and wins. The Right option $\uparrow$ is not reversible.

The Left option $\uparrow$ may be replaced by all the Left options of $\uparrow^R=*$, i.e., by 0.

Is $*\leq H$? Yes: if Right starts in $H-*=H+*$, he goes to $\uparrow+*$ or H and Left wins by playing to $\uparrow$ in either case.

Figure 15. Examining $\{\uparrow\mid\uparrow\}$ for reversibility.

For another example, consider the game $H=\{\uparrow\mid\uparrow\}$ and examine each option for reversibility (fig. 15). Check that H satisfies the *upstart equality* (ONAG, p. 77, WW, p. 73),

$$\{0\mid\uparrow\}=\uparrow+\uparrow+*=\Uparrow*\quad\text{("doubleup star").}$$

14. Impartial games (ONAG, pp. 112–130, WW, pp. 81–116)

Remember that an *impartial* game is one where the set of Left options is the same as the set of Right options. The impartial games that we have seen to far are

$$\{\ \mid\ \}=*0=0\,,\qquad \{0\mid 0\}=*1=*\quad\text{and}\quad\{0,*\mid 0,*\}=*2\,.$$

On day n, the game

$$*n = \{*0, *1, *2, \ldots, *(n-1) \mid *0, *1, *2, \ldots, *(n-1)\}$$

is born. In fact any game of the type

$$\{*a, *b, *c, \ldots \mid *a, *b, *c, \ldots\}$$

has value $*m$, where $m = \operatorname{mex}\{a, b, c, \ldots\}$, the least non-negative integer *not* in the set $\{a, b, c, \ldots\}$. To see this, notice that any option $*n$ with $n > m$ is reversible, because the option $*m$ of $*n$ is $\geqslant *m$ and so $*n$ may be deleted in favor of the options of $*m$, namely $0, *, *2, \ldots, *(m-1)$.

This is the inductive step which proves the Sprague–Grundy theorem (Sprague 1935–36, Grundy 1939), which states that every (position in an) impartial game is equivalent to a *nim-heap*.

As the Left and Right options are the same, $*n$ is its own negative, $*n + *n = 0$. Also, we need only write one set of options, and may define the *nimber*

$$*n = \{0, *, *2, \ldots, *(n-1)\}\ .$$

This exactly parallels von Neumann's (1923) definition of ordinal numbers.

Recall that the game of Nim is played with several heaps of beans. A move is to select a heap, and to remove any positive number of beans from it, possibly the whole heap. Any position in Nim is therefore the sum of several one-heap Nim games. The value of a single heap of n beans is $*n$.

15. Nim-addition (ONAG, pp. 50–51, WW, pp. 60–61)

We know that $*n + 0 = *n$ and $*n + *n = 0$. Let us calculate $*2 + * = \{0, *\} + \{0\} = \{0 + *, * + *, *2 + 0\} = \{*, 0, *2\} = *3$. Add $*$, or $*2$, to each side and obtain $*2 = *3 + *$ and $* = *3 + *2$. In general,

$$\begin{aligned} *a + *b &= \{0, *, *2, \ldots, *(a-1)\} + \{0, *, *2, \ldots, *(b-1)\} \\ &= \{0 + *b, * + *b, \ldots, *(a-1) + *b, *a + 0, *a + *, \ldots, *a + *(b-1)\} \end{aligned}$$

and we can build a nim-addition table (table 1) by noting that the options of an entry are just the earlier entries in the same row and the earlier entries in the same column. Each entry in table 1 is the least non-negative integer not appearing as an earlier entry in the same row or column. For instance, $*5 + *6 = *3$, because 3 is the first number not in the set $\{5, 4, 7, 6, 1, 0, 6, 7, 4, 5, 2\}$, i.e., the first six entries in row 5 and the first five entries in column 6. In the usual language, 3 is the *nim-sum* of 5 and 6, which is sometimes written $5 \overset{*}{+} 6 = 3$.

Contrast the two equations $*5 + *6 = *3$ and $5 \overset{*}{+} 6 = 3$. In the first, the summands are nimbers, i.e., values of impartial games, and the addition is a game sum. In the second, the summands are *nim-values* and the addition is nim-addition.

Nim-addition is perhaps better known as addition without carry in base 2,

Table 1
Nim-addition table. The stars have been omitted, i.e., the entries are nim-values instead of nimbers

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1	0	3	2	5	4	7	6	9	8	11	10	13	12	15	14
2	2	3	0	1	6	7	4	5	10	11	8	9	14	15	12	13
3	3	2	1	0	7	6	5	4	11	10	9	8	15	14	13	12
4	4	5	6	7	0	1	2	3	12	13	14	15	8	9	10	11
5	5	4	7	6	1	0	3	2	13	12	15	14	9	8	11	10
6	6	7	4	5	2	3	0	1	14	15	12	13	10	11	8	9
7	7	6	5	4	3	2	1	0	15	14	13	12	11	10	9	8
8	8	9	10	11	12	13	14	15	0	1	2	3	4	5	6	7
9	9	8	11	10	13	12	15	14	1	0	3	2	5	4	7	6
10	10	11	8	9	14	15	12	13	2	3	0	1	6	7	4	5
11	11	10	9	8	15	14	13	12	3	2	1	0	7	6	5	4
12	12	13	14	15	8	9	10	11	4	5	6	7	0	1	2	3
13	13	12	15	14	9	8	11	10	5	4	7	6	1	0	3	2
14	14	15	12	13	10	11	8	9	6	7	4	5	2	3	0	1
15	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0

vector or coordinatewise addition over GF(2), and XOR (exclusive or): it is reassuring that it also follows from the more general idea of game sum.

To summarize the Sprague–Grundy theory: the nim-value of the sum of two impartial games is the nim-sum of their separate nim-values. Impartial games belong to one of only two outcome classes: all positions are either

$\mathcal{P}$-positions	previous-player-winning	nim-value zero, or
$\mathcal{N}$-positions	next-player-winning	nonzero nim-value.

In the literature, $\mathcal{P}$-positions are sometimes called "safe" or "good" or "winning" without indicating which player enjoys this happy situation.

Table 2
Nim-sequences and periods for subtraction games

Subtraction game	Nim-sequence	(ultimate) Period
$S(1)$	$\dot{0}\dot{1}01010101\ldots$	2
$S(2)$	$\dot{0}01\dot{1}0011001100\ldots$	4
$S(3)$	$\dot{0}0011\dot{1}0001110001110\ldots$	6
$S(1,2)$	$\dot{0}1\dot{2}0120120120\ldots$	3
$S(1,2,3)$	$\dot{0}12\dot{3}0123012301230\ldots$	4
$S(1,2,3,4)$	$\dot{0}123\dot{4}01234012340\ldots$	5
$S(2,4,7)$	$0011220 3\dot{1}0\dot{2}10210210\ldots$	3
$S(2,5,6)$	$\dot{0}011021302\dot{1}0011021302100\ldots$	11
$S(4,10,12)$	$\dot{0}00011110022113300221\dot{1}0000\ldots$	22
$S(1,2,3,4,\ldots)$	$0123456789\ldots$	–

16. Subtraction games (WW, pp. 83–86 and 487–498)

These are very simple examples of impartial games, played, like Nim, with heaps of beans. A move in the game $S(s_1, s_2, s_3, \ldots)$ is to take a number of beans from a heap, provided that number is a member of the *subtraction-set*, $\{s_1, s_2, s_3, \ldots\}$. Analysis of these and other heap games is conveniently recorded by a *nim-sequence*:

$$n_0 n_1 n_2 n_3 \cdots,$$

meaning that the nim-value of a heap of h beans is n_h, $h = 0, 1, 2, \ldots$. In this section, and often later, to avoid printing stars, we say that the nim-value of a position is n, meaning that its value is the nimber $*n$.

Table 2 shows some examples: the first is a manifestation of She-Loves-Me-She-Loves-Me-Not; the last is Nim. If the subtraction-set is finite, the nim-sequence is (ultimately) periodic. But little is known about the length of the period vis-à-vis the membership of the subtraction set.

In subtraction games the nim-values 0 and 1 are remarkably related by Ferguson's Pairing Property (Ferguson 1974, WW, pp. 86 and 422): if s_1 is the least member of the subtraction-set, then

$$\mathscr{G}(n) = 1 \quad \text{just if } \mathscr{G}(n - s_1) = 0.$$

Here and later, $\mathscr{G}(n) = v$ means that the nim-value of a heap of n beans is v.

17. Take-and-break games (WW, pp. 81–106)

Guy and Smith (1956) devised a code classifying a broad range of impartial games played with heaps or rows. If the binary expansion of the kth *code digit*, d_k, in the game with code

$$\boldsymbol{d}_0 \cdot \boldsymbol{d}_1 \boldsymbol{d}_2 \boldsymbol{d}_3 \cdots$$

is

$$\boldsymbol{d}_k = 2^{a_k} + 2^{b_k} + 2^{c_k} + \cdots,$$

where $0 \leq a_k < b_k < c_k < \cdots$, then it is legal to remove k beans from a heap, provided that the rest of the heap is left in exactly a_k or b_k or c_k or $\cdots$ non-empty heaps.

In order that the game should satisfy the ending condition, $\boldsymbol{d}_0$ must be divisible by 4, i.e., $a_0 \geq 2$.

Subtraction games are the special case $\boldsymbol{d}_s = \mathbf{3}$ when s is in the subtraction-set, and $\boldsymbol{d}_k = \mathbf{0}$ otherwise.

Octal games are those with code digits $\boldsymbol{d}_k \leq \mathbf{7}$ for all k. Guy and Smith showed that an octal game is *ultimately periodic* with period p, i.e., $\mathcal{G}(n+p) = \mathcal{G}(n)$ for all $n > n_0 = 2e + p + t$, provided that $\mathcal{G}(n+p) = \mathcal{G}(n)$ for $n \leq n_0$ apart from some exceptional values of n, of which e is the largest, and $\boldsymbol{d}_k = \mathbf{0}$ for $k > t$, i.e., the maximum number of beans that may be taken from a heap is t. Whether all such finite octal games are ultimately periodic remains a difficult open equation. They cannot be *arithmetically periodic*: that is, there is no period p and *saltus* $s > 0$, such that $\mathcal{G}(n+p) = \mathcal{G}(n) + s$ for all large enough n (WW, p. 114).

Table 3 exhibits a dozen specimen games, of which the last three are *hexadecimal games* with $d_k \leq \mathbf{15} = \mathbf{F}$. Such games may be arithmetically periodic.

Recently, Gangolli and Plambeck (1989) have established the ultimate periodicity of four octal games which were previously unknown. The game $\cdot\mathbf{16}$ has period 149459 (a prime!), the last exceptional value being $\mathcal{G}(105350) = 16$. The game $\cdot\mathbf{56}$ has period 144 and last exceptional value $\mathcal{G}(326639) = 26$. The games $\cdot\mathbf{127}$ and $\cdot\mathbf{376}$ each have period 4 (with cycles of values 4, 7, 2, 1 and 17, 33, 16, 32) and last exceptional values $\mathcal{G}(46577) = 11$ and $\mathcal{G}(2268247) = 42$, respectively.

Grundy's Game (Grundy 1939, WW, p. 111), split a heap into two *unequal* heaps, continues to defy complete analysis, despite Mike Guy's calculation of the first ten million nim-values. Among these,

$$0, 1, 6, 7, 10, 11, 12, 13, 18, 19, 20, 21, 24, \ldots$$

occur quite rarely. When written in binary, these values contain an even number of ones after deleting the last digit. These rare values form a closed space (the *sparse space*) under nim-addition, while the complement forms the *common coset*:

$$\begin{aligned} \text{rare} \overset{*}{+} \text{rare} &= \text{rare} &&= \text{common} \overset{*}{+} \text{common}, \\ \text{rare} \overset{*}{+} \text{common} &= \text{common} &&= \text{common} \overset{*}{+} \text{rare}. \end{aligned}$$

If the nim-values in a sequence begin to cluster in a suitable common coset, this clustering is likely to persist. In Kayles the rare and common values are *evil* and *odious* numbers, respectively, with an even and odd number of ones in their binary expansions. On the other hand, Dawson's Kayles does not exhibit this

Table 3
Some sample take-and-break games

Code	Game	Nim-sequence
·77	*Kayles*. Knock down 1 skittle, or 2 contiguous skittles, from a row (Dudeney 1907, Loyd 1914).	Ultimate period $p = 12$, $\dot{4}1291472182\dot{7}$ except for $n = 0$, 3, 6, 9, 11, 15, 18, 21, 22, 28, 34, 39, 57, 70, nim-value is 0, 3, 3, 4, 6, 7, 3, 4, 6, 5, 6, 3, 4, 6, respectively.
·137	*Dawson's Chess*. $3 \times n$ board. White and Black pawns on ranks 1 and 3. Capturing obligatory. Looks partizan but is not (Dawson 1934, 1935).	$\dot{8}11203110332244559330113021104537\dot{4}$ except 0 for $n = 0$, 13, 34 and 2 for $n = 16$, 17, 51. $p = 34$.
·07	*Dawson's Kayles*. Knock down 2 skittles, but only if they are contiguous.	As for **·137**, but shifted one term: 0011203 . . . in place of 011203 . . . .
·6	*Officers*. Take 1 counter from any *longer* row (Blanche Descartes 1953).	No period found. $\mathscr{G}(10342) = 256$.
·007	*Treblecross*. One-dimensional tic-tac-toe (WW, pp. 93–94).	No pattern yet found.
·077	*Duplicate Kayles*. Knock down 2 or 3 contiguous skittles (Guy and Smith 1956).	$p = 24$. Kayles with each nim-value repeated, 00112233114433 . . . .
·7777	*Double Kayles*. Take up to 4 beans from a heap; leave rest in at most 2 heaps. (Guy and Smith 1956, WW, p. 98).	$p = 24$. Kayles with each nim-value g replaced by the pair $2g$, $2g + 1$ or $2g + 1$, $2g$ (according to a certain rule), 0123456732897654328945 . . . .
·156	See Kenyon (1967a,b).	$p = 349$.
·165	See Austin (1976).	$p = 1550$.
·8	(First cousin of) *Triplicate Nim*. Take 1 from heap, rest left in exactly 3 non-empty heaps.	Arithmetically periodic, $p = 3$, saltus = 1. 0000111222333444 . . . .
·3F	(**F = 15**) *Kenyon's Game*. Take 1 from heap or take 2 and leave rest in any number of heaps up to 3 (Kenyon 1967a,b).	$p = 6$, $s = 3$. 0120123453456786789 . . . .
·E	(**E = 14**) Take 1, leave rest in just 1, 2 or 3 heaps.	001234153215826514 . . . . $\mathscr{G}(246) = 128$. No known pattern.

sparse-space phenomenon. In Grundy's Game only 1273 rare values have appeared; the only one in the range $36184 < n \leq 10^7$ is $\mathscr{G}(82860) = 108$. If the rare values have indeed died out, then Grundy's Game will ultimately be periodic, but the period may be astronomical.

Amongst the comparative chaos, John Conway and Mike Guy have noted a remarkable structure in the nim-values for Grundy's Game, related to the number

59. The probability that $\mathcal{G}(n+d)=\mathcal{G}(n)$ is often as high as $\frac{1}{4}$

$$\text{if } d \text{ is } \textit{near } 59k \text{ and } d \equiv k \bmod 3 .$$

Examples of these pseudo-periods are 58, 61, 116, 119, 122, 290, 293, 296, 360, 412, 580, 583, 586, 589, 647, 650, 882, 952 and 1172, where the last four correspond to $k = 11$, 15, 16 and 20.

18. Green Hackenbush (ONAG, pp. 165–172, WW, pp. 183–190)

This is played on a picture, as is Blue–Red Hackenbush, but now all the edges are grEen, and may be chopped by Either player, making it an impartial game. Every Green Hackenbush picture has a nim-value: for example (fig. 18) the value of a string of six green edges is *6. We will see how to evaluate Green Hackenbush trees by the Colon Principle and how to reduce any picture to a forest by the Fusion Principle.

Green Hackenbush trees are examples of the *ordinal sum* (WW, p. 214) which can be defined for any two games:

$$G:H=\{G^L, G:H^L \mid G^R, G:H^R\} ,$$

where any move in G annihilates H, while a move in H leaves G unaffected. The *Colon Principle* (WW, pp. 184–185) states that $H \geqslant K$ implies $G:H \geqslant G:K$, and in particular, that $H=K$ implies $G:H=G:K$. That is, $G:H$ depends only on the *value* of H and not on its *form*. It *may* depend on the *form* of G, because there are games $G_1 = G_2$ for which $G_1:H \neq G_2:H$.

The Colon Principle applies at branch points of Green Hackenbush trees, allowing us to do nim-addition up in the air. For example, in fig. 16 at a, $*3 + *2 = *$; at b, $* + * = 0$; and at c, $* + *2 = *3$, so the value is the same as that of fig. 17, where, at d, $*2 + *2 + * + *4 = *5$, and the tree is worth *6. Notice the interplay of ordinary addition along strings, with nim-addition at branch points.

Green Hackenbush pictures involving circuits can be evaluated by the *Fusion Principle* (WW, pp. 186–188): the value of such a picture is unaltered if you identify the nodes of a circuit. The edges of the circuit then become loops, which

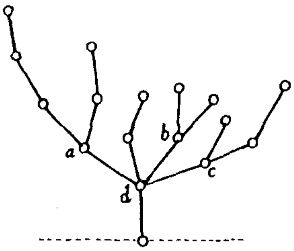


Figure 16.

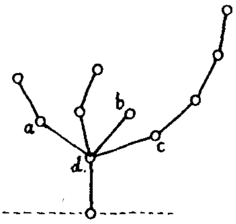


Figure 17.

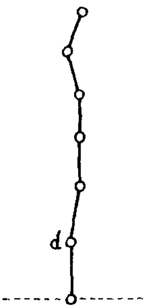
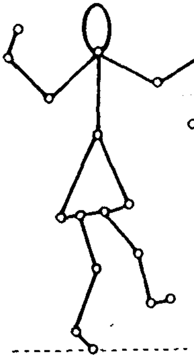
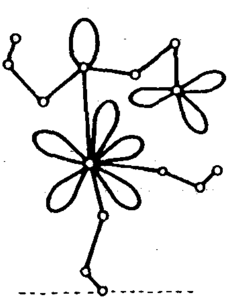


Figure 18.



=



=

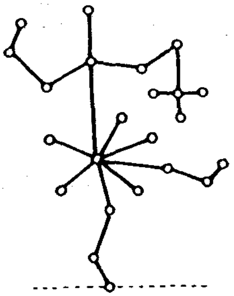


Figure 19. Figure 20. Figure 21.

may be replaced by twigs (compare figs. 20 and 21). Check that the value of fig. 19 is *8. In this way, every component of a Green Hackenbush picture can be reduced to a tree, and hence to a string, and the strings are combined by nim-addition.

19. Welter's Game (Sprague 1947, Welter 1952, 1954, Berlekamp 1972, ONAG, pp. 153–165, WW, pp. 472–481)

This is another game whose analysis involves the interplay of nim-addition and ordinary addition. It is a form of Nim with unequal heaps, but in order to keep track of empty heaps, only one of which is allowed, it is better to play it with coins on a strip of squares, numbered 0, 1, 2, . . . , with at most one coin on a square. A move is to shift a coin leftwards to any unoccupied square, possibly passing over other coins. The game ends when the k coins are on the left-most squares 0, 1, . . . , $k-1$. Figure 22 shows a position with $k = 7$.

0	●	●	●	4	●	6	7	●	9	10	11	12	●	14	15	16	17	18	19	20	●	22	23	24

Figure 22. The position {1, 2, 3, 5, 8, 13, 21} in Welter's Game.

To calculate the nim-value, or *Welter function*, $[a\,|\,b\,|\,c\,|\ldots]_k$ of the position with k coins on squares $a, b, c, \ldots,$ first note that for $k = 1$, $[a] = a$, and that for $k = 2$, $[a\,|\,b]$ is one less than the nim-sum of a and b: e.g., $[1\,|\,3] = 1$, $[5\,|\,6] = 2$. For more than two coins, *mate* the pair that is congruent modulo the highest power of 2 (it does not matter that this pair may not be unique). Remove the mated pair and find the best mated pair among the remaining $k-2$ coins. Continue until all coins are mated, except, when k is odd, for one coin, the *spinster*, s. Then, if (a, b), $(c, d), \ldots$ are the mates, $[a\,|\,b\,|\,c\,|\,d\,|\ldots]$ is the *nim-sum*

$$[a\,|\,b] \overset{*}{+} [c\,|\,d] \overset{*}{+} \cdots (\overset{*}{+} [s]),$$

where the last term is included just if k is odd.

In fig. 22 the best mates are (5, 21), then (1, 13), then (2, 8), and 3 is the spinster, so the nim-value is

$$\begin{aligned}[1\,|\,2\,|\,3\,|\,5\,|\,8\,|\,13\,|\,21] &= [5\,|\,21] \overset{*}{+} [1\,|\,13] \overset{*}{+} [2\,|\,8] \overset{*}{+} 3\\ &= 15 \overset{*}{+} 11 \overset{*}{+} 9 \overset{*}{+} 3 = 14\,.\end{aligned}$$

It turns out that $[a\,|\,b\,|\,c\,|\,d] = 0$ just if the nim-sum $a \overset{*}{+} b \overset{*}{+} c \overset{*}{+} d = 0$, so Welter's Game with four coins can be played with a Nim-like strategy. To play with three coins, imagine a fourth coin on an extra square -1, and add 1 to the numbers labelling the squares while you calculate your move. For example, $\{2, 5, 8\}$ is like $\{0,3,6,9\}$, where the winning move would be to $\{0,3,5,6\}$, so move to $\{2,4,5\}$.

The mating method makes it easy to calculate the nim-value of a Welter position, but it is not so easy to find the good moves which make the nim-value 0. However, there is a remarkable connexion with *frieze patterns* (Conway and Coxeter 1973, WW, pp. 475–480), which work for nim-addition as well as for multiplication and ordinary addition, and which allow you (or your computer) both to calculate the value of the Welter function and to invert it.

Start with a row of zeros above the Welter position that you want to evaluate, and manufacture a frieze pattern (so called because, when it is extended to the right, it eventually repeats periodically) by completing diamonds,

$$\begin{array}{ccc} & b & \\ a & & d \\ & c & \end{array} \quad \text{using the rule} \quad a \overset{*}{+} d = (b \overset{*}{+} c) + 1\,,$$

so that $c = [a\,|\,d] \overset{*}{+} b$, where the sums are still nim-sums. Lo and behold (fig. 23) the value of the Welter function appears at the foot of the pattern, as follows from a formula on page 159 of ONAG.

```
0   0   0   0   0   0   0   0
  1   2   3   5   8   13  21
    2   0   5   12  4   23
      3   7   13  15  31
        3   12  13  11
          9   13  10
            15  11
              14
```

Figure 23. Calculating the Welter function from a frieze pattern.

If you want to change the value $n = [a\,|\,b\,|\,c\,|\,\ldots]$ to some $n' \neq n$, then there are unique $a' \neq a$, $b' \neq b$, $c' \neq c, \ldots$ such that

$$[a'\,|\,b\,|\,c\,|\,\ldots] = n' = [a\,|\,b'\,|\,c\,|\,\ldots] = [a\,|\,b\,|\,c'\,|\,\ldots] = \cdots$$

Then $[a\,|\,b\,|\,c\,|\,\ldots] = n$ remains true if we replace any *even* number of a, b, $\ldots, n$ by the corresponding primed letters. This *Even Alteration Theorem*

(ONAG, pp. 160–162; WW, p. 477) may be written

$$\left[\begin{array}{c|c|c|c} a & b & c & \\ a' & b' & c' & \cdots \end{array}\right] = \begin{array}{c} n \\ n' \end{array}$$

To find a', b', c', . . . corresponding to a given n', continue the bottom row of the frieze pattern, n, n', n, n', n, . . . alternately, then extend the pattern to the right, using the same diamond rule. You will discover that the defining row, a, b, c, . . . continues a', b', c', ...

In fig. 24 we find the good moves in the position $\{1, 2, 3, 5, 8, 13, 21\}$ by choosing $n' = 0$ and extending the pattern of fig. 23. If you extend it even further to the right, you will see why it is called a frieze pattern. If you believe the algorithm, and read the second row of fig. 24,

$$\left[\begin{array}{c|c|c|c|c|c|c} 1 & 2 & 3 & 5 & 8 & 13 & 21 \\ 15 & 0 & 37 & 35 & 10 & 11 & 19 \end{array}\right] = \begin{array}{c} 14 \\ 0 \end{array}$$

```
 0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
   1   2   3   5   8  13  21  15   0  37  35  10  11  19
     2   0   5  12   4  23  25  14  36   5  40   0  23
       3   7  13  15  31  24  25  41   5  15  45  29
         3  12  13  11  17  25  33  15  12   9  47
           9  13  10   6  31  46   4   7  11   8
            15  11   0   9  41   8  13   7  11
              14   0  14   0  14   0  14   0
```

Figure 24. Inverting the Welter function using a frieze pattern.

Check that each move leads to a $\mathscr{P}$-position. Some of the suggested moves, e.g., 1 to 15, 3 to 37, are not legal, but, provided $n' < n$, you will always find one that is; in fact there is always an odd number of legal good moves. Here there are three good moves:

2 to 0, 13 to 11, and 21 to 19.

We can even give you a strategy for the misère form (last player losing) of Welter's Game, if your are willing to learn about *Abacus Positions* (WW, pp. 478–481).

20. Coin-turning games (WW, pp. 429–456)

Several of the impartial games we have already mentioned, and a wide range of new games, can be realized by an idea of Hendrik Lenstra. *Turning Turtles* was originally played with turtles, but it is less cruel to play it with a row of coins (fig. 25). A move is to turn a head to a tail, with the additional option of turning at most one other coin, to the left of it. This second coin may go from head to tail, or from tail to head. The game is over when all coins show tails, and the last player wins.

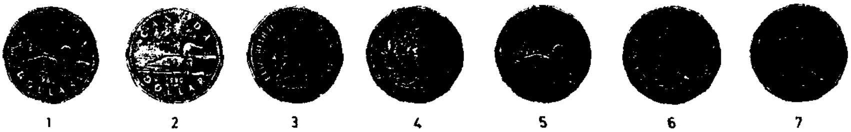


Figure 25. A Turning Turtles position, with coins 3, 4, 6, and 7 showing heads.

We leave you to verify that this is a disguise for Nim: if you number the coins 1, 2, 3, . . . from the left, then the value of coin n is $*n$ if it is a head, 0 if it is a tail. The value of a general position is the nim-sum of the heads. For example, the good moves in fig. 25 are to turn coin 6 to a tail; or to turn 7 to a tail and 1 to a head; or to turn 4 to a tail and 2 to a head.

Mock Turtles is played in the same way, but a move may turn one, two or three coins, provided the rightmost turned goes from head to tail (this is to make the game satisfy the ending condition). We now number the coins from *zero* (the Mock Turtle) and find the nim-value (or Grundy function), $\mathcal{G}(n)$, of the nth coin, when head up, to be

$$\begin{array}{lllllllllllllllllllll} n = & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 & 17 & 18 & \ldots, \\ \mathcal{G}(n) = & 1 & 2 & 4 & 7 & 8 & 11 & 13 & 14 & 16 & 19 & 21 & 22 & 25 & 26 & 28 & 31 & 32 & 35 & 37 & \ldots . \end{array}$$

These are the *odious numbers* which we met as common values in Kayles.

$$\mathcal{G}(n) = 2n \text{ or } 2n + 1 .$$

To find which, write n in binary and append a check digit, 0 or 1, to make an *odd* number of 1-digits.

Moebius, *Mogul*, and *Moidores* are the corresponding games in which a move turns up to t coins, where $t = 5$, 7, and 9, respectively. We consider only odd values of t, because the *Mock Turtle Theorem* gives us the results for even values of t:

> Every nim-value for the $t = 2m + 1$ game
> is an odious number.
> The corresponding value for the $t = 2m$ game
> is got by dropping the final binary digit.

The nim-values for coins 0 to 17 (when head-up) in Moebius are shown in table 4. The structure of the $\mathcal{P}$-positions in 18-coin Moebius is revealed on replacing the coin numbers by the labels in the third row.

Table 4
Eighteen-coin Moebius gives the game its name

Coin number	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Nim-value	1	2	4	8	16	31	32	64	103	128	171	213	256	301	342	439	475	494
Label	∞	1	4	0	−4	−1	5	6	−8	2	−3	−5	8	3	−7	7	−6	−2

Coins 0 to 5, with labels ∞, 0, ± 1, ± 4, clearly form a $\mathcal{P}$- position (whichever ones you turn over, I will turn over the rest). Starting from this, or any $\mathcal{P}$-position, we can find others by operating on the labels with any *Möbius transformation* (modulo 17):

$$x \to \frac{ax+b}{cx+d} \quad \text{with } ad - bc = 1 \, .$$

There are	1 +	102 +	153 +	153 +	102 +	1	$\mathcal{P}$-positions
with	0	6	8	10	12	18	heads, respectively .

They correspond to the 2^9 codewords in the [18, 9, 6] extended quadratic residue code. If we drop the Mock Turtle (at ∞) we have the $t = 4$ game on 17 coins, whose $\mathcal{P}$-positions correspond to codewords in the [17, 9, 5] quadratic residue code.

Similarly if we play 24-coin Mogul ($t = 7$, turn up to 7 coins) we find

	1 +	759 +	2576 +	759 +	1	$\mathcal{P}$-positions
with	0	8	12	16	24	heads, respectively ,

coinciding with the 2^{12} codewords of the extended [24, 12, 8] Golay code. With $t = 6$ and 23 coins we have the perfect [23, 12, 7] Golay code (Curtis 1976, 1977).

In the *Ruler Game* any number of *contiguous* coins may be turned (rightmost always going from head to tail). If the coins are numbered from 1, the nim-value, $\mathcal{G}(n)$, is the highest power of 2 that divides n.

In *Turnips* (or *Ternups*) a move turns three equally spaced coins. Number the coins from 0 and write n in *ternary* (base 3). $\mathcal{G}(n)$ is the kth odious number if the last 2-digit in the ternary expansion is in the kth place from the right, or $\mathcal{G}(n) = 0$ if there is no 2-digit.

There is a plethora of such coin-turning games. They can also be played on a two-dimensional array of coins. For example, we can play the Cartesian product, $A \times B$, of two one-dimensional games, in which a move is to turn all coins with coordinates (a_i, b_j), where $\{a_i\}$ and $\{b_j\}$ are sets of coins constituting legal moves in games A and B, respectively. To satisfy the ending condition, the "most northeasterly" coin turned must go from head to tail (fig. 26).

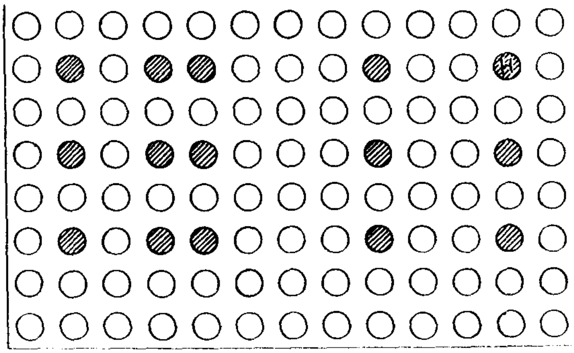


Figure 26. A typical move in Moebius × Turnips.

The nim-value of a (head-up) coin in such a game is given by the *Tartan Theorem*:

The nim-values for the game $A \times B$ are the nim products of those for A and B:

$$\mathcal{G}_{A+B}(a,b) = \mathcal{G}_A(a) \overset{*}{\times} \mathcal{G}_B(b)\,,$$

where $\mathcal{G}_A(a)$ is the nim-value of coin number a in game A, etc. and $\overset{*}{\times}$ denotes nim-multiplication. Nim-multiplication (ONAG, pp. 52–53, Lenstra 1977) may be defined from the field laws (e.g., associativity and distributivity over nim-addition), together with the rule

$$\boxed{\begin{gathered} n \overset{*}{\times} N = n \times N \quad \text{for } n < N\,, \\ N \overset{*}{\times} N = 3N/2\,, \end{gathered}}$$

where N is any *Fermat power* of 2 $(2, 4, 16, \ldots, 2^{2^h}, \ldots)$. For example, $2 \overset{*}{\times} 2 = 3$, because 2 is a Fermat power, while $2 \overset{*}{\times} 3 = 2 \overset{*}{\times} (2 \overset{*}{+} 1) = 3 \overset{*}{+} 2 = 1$. And

$$\begin{aligned}(13 \overset{*}{\times} 7) &= (4 \overset{*}{\times} 3 \overset{*}{+} 1) \overset{*}{\times} 7 = (4 \overset{*}{\times} (2 \overset{*}{+} 1)) \overset{*}{\times} (4 \overset{*}{+} 2 \overset{*}{+} 1) \overset{*}{+} 7 \\ &= 4 \overset{*}{\times} 4 \overset{*}{\times} (2 \overset{*}{+} 1) \overset{*}{+} 4 \overset{*}{\times} 2 \overset{*}{\times} (2 \overset{*}{+} 1) \overset{*}{+} 4 \overset{*}{\times} (2 \overset{*}{+} 1) \overset{*}{+} 7 \\ &= 6 \overset{*}{\times} (2 \overset{*}{+} 1) \overset{*}{+} 4 \overset{*}{\times} (3 \overset{*}{+} 2) \overset{*}{+} 4 \overset{*}{\times} 3 \overset{*}{+} 7 \\ &= (4 \overset{*}{+} 2) \overset{*}{\times} (2 \overset{*}{+} 1) \overset{*}{+} 4 \overset{*}{\times} 1 \overset{*}{+} 12 \overset{*}{+} 7 \\ &= 4 \overset{*}{\times} 2 \overset{*}{+} 2 \overset{*}{\times} 2 \overset{*}{+} 4 \overset{*}{+} 2 \overset{*}{+} 4 \overset{*}{+} 11 = 8 \overset{*}{+} 3 \overset{*}{+} 9 = 2\,. \end{aligned}$$

The assiduous reader will verify that the nim cube roots of 1 are 1, 2, and 3, and the nim fifth roots are 1, 8, 10, 13, 14.

21. Lexicodes

Conway and Sloane (1986) have noticed a striking connexion between the calculation of nim-values, involving, as it does, the mex of a set of non-negative integers, which is the lexicographically first number not in the set; and the construction of error-correcting codes by successively choosing the lexicographically first codeword to satisfy the required distance, weight, and other conditions. Both processes use the greedy algorithm, while leads to some not-always-expected isomorphisms.

Unrestricted binary lexicodes are linear (Marc Best 1975, unpublished). If the base is a power of 2, unrestricted lexicodes are closed under nim-addition. If the

Table 5
Close connexions between coin-turning games and lexicodes

Games (WW, chapter 14)	Codes (Conway and Sloane 1986)
Turning Turtles on 7 coins; Nim with heaps of up to 7 beans.	The [7, 4, 3] Hamming code.
Mock Turtles on 8 coins.	The [8, 4, 4] extended Hamming code.
The Mock Turtle Theorem.	Extending codes by a parity check digit.
Moebius on 17 (18) coins	The (extended) [17, 9, 5] ([18, 9, 6]) quadratic residue code.
Mogul on 23 (24) coins; automorphism group M_{23} (M_{24}); the MOG (Curtis 1976, 1977).	The (extended) [23, 12, 7] ([24, 12, 8]) Golay code. (The Steiner system $S(5, 8, 24)$. The Mathieu groups M_{23}, M_{24}. The Leech lattice.)
Moidores ($t = 9$) on 27 (31) coins.	Binary [27, 9, 10] ([31, 12, 10]) lexicode.
Welter's Game; Nim with unequal heaps; connexion with frieze patterns.	Constant-weight binary lexicodes with distance 4.
Mathematical Blackjack; Mathieu's Vingt-et-un (Curtis 1984).	The [12, 4] lexicodes of constant weight 6 with Ryba's restriction (sum of digits at least 21). (Steiner system $S(5, 6, 12)$. The Mathieu group M_{12}.)

base is a Fermat power of 2, i.e., 2^{2^h}, then unrestricted lexicodes are also closed under nim-multiplication. Table 5 gives a glimpse of the numerous connexions between impartial games and error-correcting codes, and with several other branches of combinatorics.

22. The remoteness function (WW, pp. 259–266)

Steinhaus (1925) gave an early analysis of impartial games, using what Smith (1966) has since called the *remoteness function*. This is useful for games where the idea of sum does not arise naturally, or does not arise at all (if a move may affect more than one component). We will later see how it serves to analyse *joins* of games in which moves must simultaneously be made in *all* components. It can apply to partizan games (where we define Left and Right remotenesses) and in places where we may want to amend the rules: for example, in misère play (where the last player *loses*), or where the ending condition does not hold (so that some remotenesses may be infinite).

Intuitively, the remoteness is the number of turns that the game lasts when the winner is trying to win as quickly as possible, while the loser tries to postpone defeat as long as possible. In normal play the last player wins, so she wants to move to a terminal position (remoteness 0), and, more generally, to a position of *even* remoteness, forcing her opponent always to move to positions of odd remoteness. Terminal positions have remoteness 0; positions with an option of remoteness 0 have remoteness 1; positions *all* of whose options have remoteness 1, are of remoteness 2; and so on.

Rules for calculating the remoteness

If there is an option of even remoteness:	1 + least even remoteness	$\mathcal{N}$-position;
If all options have odd remoteness:	1 + greatest odd remoteness	$\mathcal{P}$-position;
If there are no options:	0	terminal position.

The $\mathcal{P}$-positions are those of even remoteness; the $\mathcal{N}$-positions those of odd remoteness. If the game does not satisfy the ending condition, it may not always be possible to continue assigning remotenesses according to these rules. The remaining positions are *$\mathcal{O}$-positions* (open positions, for which neither player has a winning strategy), to which we assign *infinite* remoteness.

Epstein's (1967) *Put-or-Take-a-Square* game (WW, pp. 484–486 and 501–502) is played with a single heap of beans. A move is to add or take away the largest perfect square number of beans in the heap. The object is to take the last bean, so the empty heap has remoteness 0, and a positive perfect square number of beans has remoteness 1. Figure 27 shows some heap sizes with legal moves indicated by arrows. We can next assign remoteness 2 to heaps of 5 and 20, because *both* options have remoteness 1. Then 11, 14, 21, 30, 41, 54 have remoteness 3, since in each case there is an option of remoteness 2 (and no option of remoteness 0). A heap of 29 has remoteness 4 because both options (4 and 54) have odd remoteness, and the larger is 3.

Each of 2 and 3 is an option of the other, and sensible players will not choose the other options, 1 and 4, because these have (odd) remoteness 1 and lose (immediately). So best play goes 2, 3, 2, 3, 2, . . . and the remotenesses are infinite. Table 6 shows a few $\mathcal{P}$- and $\mathcal{N}$-positions, but the great majority of positions,

$$2, 3, 6, 7, 8, 10, 12, 13, 15, 17, 18, 19, 22, 23, \ldots,$$

are $\mathcal{O}$-positions with infinite remoteness.

The game of *Fair Shares and Varied Pairs* is played with heaps (WW, pp. 359, 390). A move either divides a heap into two or more *equal-sized* heaps or combines two *different-sized* heaps into a single heap. The terminal positions, of remoteness 0, are those with all heaps of size 1: i.e., the bottom row in table 7, where exponents indicate repetitions of heap size. They next row up shows positions with just one splittable heap (of size >1): these have remoteness 1.

A dramatic change in the game occurs when we play with 11 beans. There is one $\mathcal{P}$-position, 1^{11}, ten $\mathcal{N}$-positions, $m.\,1^{11-m}$ $(m = 2, 3, \ldots, 11)$, of remoteness 1, and the other 45 partitions of 11, all those with two or more splittable heaps, are all $\mathcal{O}$-positions.

Simon Norton's game of *Tribulations* (WW, pp. 486 and 501–502) is similar to

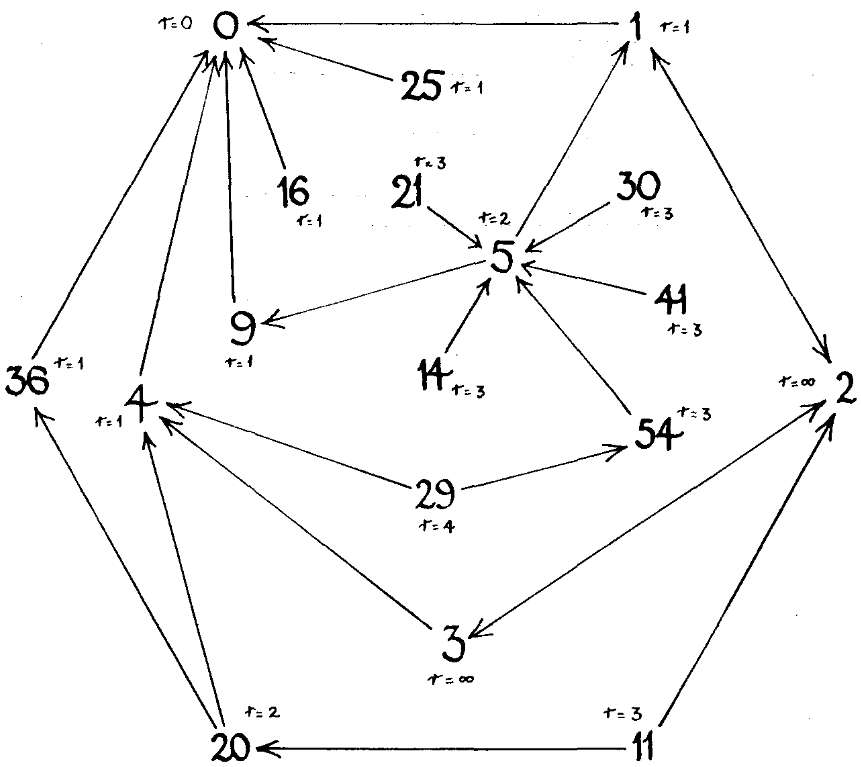


Figure 27. Assigning remotenesses in Epstein's game.

Epstein's Put-or-Take-a-Square game, but with triangular numbers, 1, 3, 6, 10, 15, . . . , used in place of squares. The largest possible triangular number is taken from or added to the heap. Norton conjectures that there are no $\mathcal{O}$-positions, and that the $\mathcal{N}$-positions outnumber the $\mathcal{P}$-positions in golden ratio, $\tau = (1 + \sqrt{5})/2 \approx 1.618$; these conjectures have been verified for heap sizes of less than 5000.

For Mike Guy's game of *Fibulations* (similar to Simon Norton's, but using the Fibonacci numbers plus one, 1, 2, 3, 4, 6, 9, 14, 22, . . .) the corresponding assertions can be proved, and indeed there is a complete theory (WW, pp. 486 and 501).

John Isbell's game of *Beanstalk* (Guy 1986) starts with a positive integer, n_0. Moves to successive positions, n_1, n_2, . . . are given by

$$n_{i+1} = 3n_i \pm 1 \quad n_i \text{ odd}, \qquad n_{i+1} = \tfrac{1}{2}n_i \quad n_i \text{ even}.$$

For John Conway's game of *Beans-Don't-Talk*, the rule is $n_{i+1} = (3n_i \pm 1)/2^*$, where 2^* is the largest power of 2 which will divide the numerator. The object in

Table 6
Remotenesses of heap sizes in Epstein's Put-or-Take-a-Square game

ıess	$\mathscr{P}$-positions	$\mathscr{N}$-positions	Remoteness
	0		
		1, 4, 9, 16, 25, 36, 49, 64, 81, 100, 121, 144, 169, 196, 225, 256, . . .	1
	5, 20, 45, 80, 145, . . .		
		11, 14, 21, 30, 41, 44, 54, 69, 86, 105, 120, 126, 141, 149, 164, 174, 189, 216, 291, . . .	3
	29, 101, 116, 135, 165, 236, . . .		
		52, 71, 84, 208, 254, 284, 296, . . .	5
	257, . . .		
		136, 436,	7
8	477, . . .		
		252, 342, . . .	9
10	173, . . .		
		92, . . .	11

Table 7
Remotenesses of partitions in Fair Shares and Varied Pairs

Remoteness	Beans										Remoteness
	1	2	3	4	5	6	7	8	9	10	
15										64	15
14									42^21	$82, 42^3$	14
13									$2^41, 32^3$	$2^5, 621^2$	13
12									2^31^3	2^41^2	12
11									$2^21^5, 3^221, 63$	$4321, 4^22, 32^31$	11
10									$321^4, 3^3$	$421^4, 4^21^2, 3^22^2$	10
9								62	3^21^3	$52^21, 2^31^4, 42^21^2, 32^21^3, 62^2$	9
8							2^31	$2^4, 421^2, 4^2$	431^2	2^21^6	8
7							2^21^3	$2^31^2, 32^21, 42^2$	$421^3, 32^21^2$	$321^5, 3^31$	7
6						42	321^2	2^21^4	521^2	3^21^4	6
5						2^3	3^21	321^3	$531, 4^21$	$3^221^2, 431^3, 521^3$	5
4						2^21^2	43	3^21^2	54	531^2	4
3					2^21	321	$32^2, 421$	$3^22, 431, 521$	$432, 52^2, 621$	$541, 532, 631, 721, 43^2$	3
2				2^2	32	3^2	52	53	72	$5^2, 73$	2
1		2	3, 21	$4, 31, 21^2$	. . .		those with just one splittable heap			. . .	1
0	1	1^2	1^3	1^4	1^5	1^6	1^7	1^8	1^9	1^{10}	0

both games is to move to 1. It is not known if there are any $\mathcal{O}$-positions in either game. The first few remotenesses are shown in table 8.

Table 8
Remotenesses for Beanstalk and Beans-Don't-Talk

$r(B)$	0	1	7	2	5	8	65	3	11	6	63	9	9	66	17	4	61	12	69	7	7	64	15	...
n_0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	...
$r(BDT)$	0	3	1	5	1	6	2	5	3	3	1	6	4	9	7	24	5	10	3	8	1	13	6	...

23. A dozen ways for simultaneous plays (Smith 1966, ONAG, pp. 173–187)

So far we have combined games by their (disjunctive) *sum*, and usually assumed normal play, with the last player winning in the last component to end. But there are other ways to play several games simultaneously.

When it is your turn to move, instead of playing in just ONE component, you might move in ALL of them, or in SOME (maybe all, but at least one). We will call these alternative combinations *joins* or *unions*, respectively, to distinguish them from sums.

How do games end? For sums and unions, it is natural to continue so long as there is a component with a legal move. For joins, it is more natural to stop as soon as play ends in *any* component, since it is no longer possible to move in *every* component. However, it is possible to consider the opposite state of affairs in each case. These possibilities are summarized in table 9, with mnemonic hints, using the initial letters of the names of the games, for the methods of analysis.

Table 9
Six ways to combine and end games

Name	Compound	Moves made in	Finish determined when	Analysis
Proper sum	Disjunctive	ONE	ALL (long)	Plain nim-value
Ouick sum	*Diminished* Disjunctive	ONE	FIRST (short)	Oueer, or foreclosed nim-value (ONAG, p. 178–179)
Rapid join	Conjunctive	ALL	FIRST (short)	Remoteness
Slow join	*Continued* Conjunctive	ALL	ALL (long)	Suspense number
Tardy union	Selective	SOME	ALL (long)	Tally (WW, p. 281) (toll and timer)
Urgent union	*Severed* Selective	SOME	FIRST (short)	Unrestricted tally
		component(s)	component(s) finished	

And who is the winner? In *normal play* the last person to play wins: you lose if you are unable to move. But in *misère play* the last player to move *loses*. These two outcomes, combined with the six possibilities of table 9, provide a dozen methods of conducting simultaneous play. We will do no more than comment briefly on a few of these. Indeed, we have no theory for misère unions of partizan games, and misère sums, for even quite simple impartial games, quickly get beyond our grasp.

24. Misère Nim and an awful warning (ONAG, pp. 136–152, WW, pp. 393–426, Grundy and Smith 1956)

Misère play of ordinary sums is very complicated, even in the impartial case. When Bouton (1901–02) analyzed (and named) Nim, he noted that his analysis also applied to the misère form.

Play always to $\mathscr{P}$-positions (nim-sum zero) until there is just one heap with more than one bean. Then take all, or all but one, of that heap, to leave an *even* number of singleton heaps in *normal* play, or an *odd* number of singletons in *misère play*.

This simple rule has misled several writers into thinking that a similar device can be used in any impartial game. This is true for very very few games. For example, the known single-heap $\mathscr{P}$-positions in Grundy's Game contain 0, 1, 2, 4, 7, 10, 20, 23, 26, 50, 53, 270, 273, 276, 282, 285, 288, 316, 334, 337, 340, 346, 359, 362, 365, 386, 389, 392, 566, 630, 633, 636, 639, 673, 676, 682, 685, 923, 926, 929, 932, or 1222 beans, and there are probably no others; those in the misère version contain 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 50, or 94 beans (Allemang 1984), and there may be others.

Complete analyses of specific misère games are only known for *tame games* (WW, pp. 407–410). In a few sporadic cases it has been possible to exhibit a winning strategy without having to give a complete analysis. Notable examples are Lucasta (WW, pp. 556–563), Welter's Game (WW, pp. 480–481), and Sibert's recent analysis (Sibert and Conway 1992) of misère Kayles.

25. Joins (WW, pp. 257–278)

We can analyze (rapid) joins, partizan or impartial, normal or misère, using Steinhaus's remoteness function. If the game is partizan, calculate the Left and Right remotenesses separately, using the remotenesses of the Right and Left options, respectively. For misère play, reverse the parity of the normal remoteness rules (table 10). For impartial joins, omit all references to Left and Right, giving a single normal (or misère) remoteness function. Since a rapid join of games ends when the *first* component ends, the remoteness (Left or Right or impartial, normal or misère) of the RAPID JOIN is the

LEAST REMOTENESS of any component.

To win a rapid join, move to a position in which your *opponent's* remoteness is

Table 10
Remoteness rules for rapid joins

In each component, to calculate the			
LEFT NORMAL	RIGHT NORMAL	LEFT MISÈRE	RIGHT MISÈRE
remoteness, take it as ZERO if there is no option for			
LEFT	RIGHT	LEFT	RIGHT
Otherwise ADD ONE to the LEAST			
EVEN RIGHT	EVEN LEFT	ODD RIGHT	ODD LEFT
remoteness of any			
LEFT	RIGHT	LEFT	RIGHT
option which has such a remoteness, else ADD ONE to the GREATEST			
ODD RIGHT	ODD LEFT	EVEN RIGHT	EVEN LEFT
remoteness of any			
LEFT	RIGHT	LEFT	RIGHT
option which has such a remoteness.			

EVEN in NORMAL PLAY, ODD in MISÈRE PLAY.

To play *slow* joins, in which play continues in the remaining components even though some have already ended, use the *suspense number* (WW, pp. 266–272), which has the opposite philosophy to that of remoteness: use cat-and-mouse tactics; if you are winning, spin the game out as long as possible; if you are losing, aim to get it over quickly. Convert table 10 to give *suspense rules* by interchanging GREATEST and LEAST throughout. The suspense of a SLOW JOIN is then the

GREATEST SUSPENSE of any component,

and, to win a slow join, again move to a position where your *opponent's* suspense number is EVEN or ODD according as play is NORMAL or MISÈRE.

We illustrate these ideas with an analysis of several variants of the game *All the King's Horses*, taken, with kind permission of Academic Press and the authors, from chapter 9 of WW. This is played as a join: each player moves *every* horse in one or other of the two ways indicated in fig. 28. There can be arbitrarily many horses on a square and they are all moved by both players. A player is unable to move if there is any one horse he cannot move. Under normal play he then loses, but under misère play he wins. Left and Right remotenesses are calculated as shown in fig. 29. Table 11 shows the Left and Right remotenesses in (a) normal play, and (b) misère play.

If we play All the King's Horses as a *slow join*, so that the *last* horse to reach home determines the outcome, we must allow a player to pass for a horse he cannot move: the game ends when *all* the horses reach home. Best play is guided by *suspense numbers*, shown in table 12.

We can also play this version as an *impartial game*: each player moves every horse whenever possible, using any of the four moves shown in fig. 28. Play is guided by a single suspense number, as in table 13.

Another variant, a compromise between the rapid join and the slow join, is to

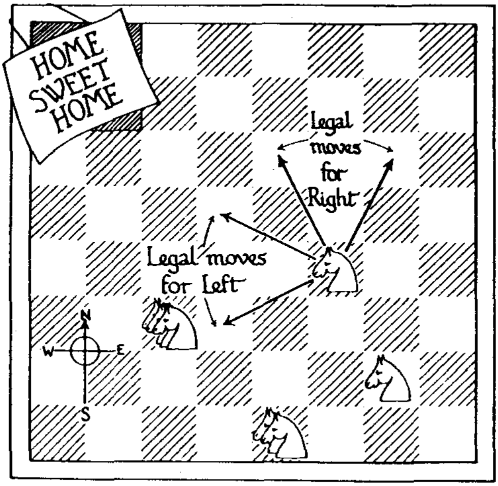


Figure 28. How horses head for home.

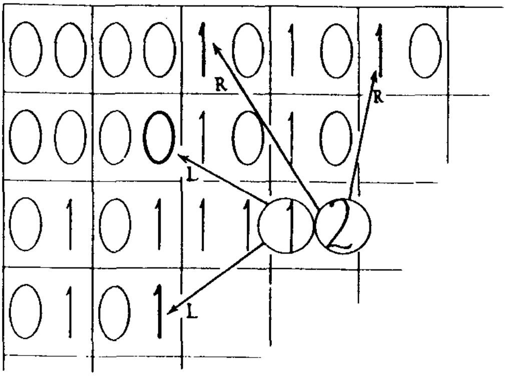


Figure 29. How remote is a horse?

allow a player to pass for a horse he cannot move, provided his opponent *can* move it. Then, the first horse *home* determines the outcome, not the first horse stuck.

In fig. 30, the misère remotenesses are being calculated. Left can move to a position of Right remoteness 3, so his remoteness is 3 + 1 = 4. Right has no move, but can pass to the same position with Left to move, so the Right remoteness is 4 + 1 = 5. Table 14 shows the remotenesses for this variant, and table 15 gives the (single) remotenesses when it is played impartially, each player having up to four possible moves for each horse.

Table 11
Left and Right remotenesses in (a) normal, and (b) misère play (A = 10, B = 11, C = 12)

00	00	10	10	10	10	10	10
00	00	10	10	10	10	10	10
01	01	11	12	12	12	12	12
01	01	21	22	22	32	32	32
01	01	21	22	22	32	32	32
01	01	21	23	23	33	34	34
01	01	21	23	23	43	44	44
01	01	21	23	23	43	44	44

(a) First horse stuck loses

00	00	10	10	10	10	10	10
00	00	20	30	30	30	30	30
01	02	22	42	52	52	52	52
01	03	24	44	64	74	74	74
01	03	25	46	66	86	96	96
01	03	25	47	68	88	A8	B8
01	03	25	47	69	8A	AA	CA
01	03	25	47	69	8B	AC	BB

(b) First horse stuck wins

Table 12
Left and Right suspense numbers for a slow join: (a) normal play, (b) misère play

00	00	12	12	34	34	56	56
00	00	12	12	34	34	56	56
21	21	11	12	32	34	54	56
21	21	21	22	22	34	34	56
43	43	23	22	22	34	34	56
43	43	43	43	43	33	34	54
65	65	45	43	43	43	44	44
65	65	65	65	65	45	44	44

(a) Last horse home wins

00	00	12	12	45	34	56	56
00	00	23	12	34	34	67	56
21	32	22	22	44	36	56	56
21	21	22	44	34	44	56	77
54	43	44	43	44	56	66	56
43	43	63	44	65	66	58	66
65	76	65	65	66	85	66	77
65	65	65	77	65	66	77	66

(b) Last horse home loses.

26. Unions (WW, pp. 279–306)

Before we analyze unions, in which you are allowed to move in any positive number of components, remind yourself of the distinction between hot and cold games, which we illustrated with the hot game of Domineering and the cold game of Blue–Red Hackenbush. The following simple example, in which Left and Right each have a single option, a number, will suffice:

If $x < y$, then the game $\{x \mid y\}$ is *cold*, i.e., a number.

Table 13
Impartial suspense numbers in (a) normal, and (b) misère play

0	0	1	1	2	2	3	3
0	0	1	1	2	2	3	3
1	1	1	3	3	3	3	3
1	1	3	3	3	3	5	5
2	2	3	3	4	4	5	5
2	2	3	3	4	4	5	5
3	3	3	5	5	5	5	5
3	3	3	5	5	5	5	6

(a) Last horse home wins

0	0	1	1	3	2	3	3
0	0	2	1	2	2	4	3
1	2	2	4	2	4	4	4
1	1	4	3	2	3	6	5
3	2	2	2	5	4	3	4
2	2	4	3	4	3	6	5
3	4	4	6	3	6	5	6
3	3	4	5	4	5	6	7

(b) Last horse home loses

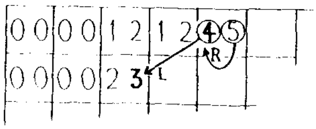


Figure 30. Right is a bit more remote because he is stuck.

Table 14
Left and Right remotenesses for a not so rapid join: (a) with normal play, (b) with misère play (A = 10, B = 11)

00	00	12	12	34	34	56	56
00	00	12	12	34	34	56	56
21	21	11	14	34	36	56	56
21	21	41	44	44	56	56	76
43	43	43	44	44	56	56	76
43	43	63	65	65	55	58	76
65	65	65	65	65	85	88	86
65	65	65	67	67	67	68	66

(a) First horse home wins

00	00	12	12	45	56	56	56
00	00	23	34	34	34	67	78
21	32	22	42	42	56	56	56
21	43	24	44	54	64	74	77
54	43	24	45	66	76	86	76
65	43	65	46	67	66	98	A8
65	76	65	47	68	89	88	B9
65	87	65	77	67	8A	9B	AA

(b) First horse home loses.

If $x > y$, then the game $\{x \mid y\}$ is *hot*.
If $x = y$, then the game $\{x \mid y\}$ is *tepid*, in fact it is $x + *$ or $x*$.

The best strategy in playing a union of several components is to move in *all* the *hot* ones, and in *none* of the cold. The first phase of the play is like a *join* of the

Table 15

Impartial remotenesses for the not so rapid join: (a) with normal play, (b) with misère play

0	0	1	1	2	2	3	3
0	0	1	1	2	2	3	3
1	1	1	1	3	3	3	3
1	1	1	3	3	3	3	5
2	2	3	3	4	4	5	5
2	2	3	3	4	4	5	5
3	3	3	3	5	5	5	5
3	3	3	5	5	5	5	6

(a) First horse home wins

0	0	1	1	3	4	3	3
0	0	2	3	2	2	4	5
1	2	2	2	2	4	4	4
1	3	2	3	4	5	6	5
3	2	2	4	5	4	5	6
4	2	4	5	4	7	6	7
3	4	4	6	5	6	7	6
3	5	4	5	6	7	6	7

(b) First horse home loses

hot components. Players will want to continue so long as there is a hot component left, so it is a *slow* join. When all components are cold, i.e., numbers, players will be reluctant to play and will move in only one component, the one where least harm is done. The union becomes like an ordinary sum.

The selective theory of unions combines
the disjunctive theory of sums with
the conjunctive theory of slow joins.

We indicate the state of play by a pair of (Left and Right) tallies (WW, pp. 280–306). A *tally* consists of a toll, and a timer, written as a subscript. If Left (or Right) starts, the Left (or Right) *toll* is the (numerical) value that the component will acquire when the hot phase is over. The Left (or Right) *timer* is the number of moves that the hot phase lasts. Think of the timers as suspense numbers (though care will be needed in their calculation for *tepid* games, see later).

To find the Left (or Right) tally for the union of several components, *add* the Left (or Right) tolls and take the *greatest* of the Left (or Right) timers.

The best *Left* option in a component is one with *greatest Right* toll, and, among those, one with *largest even* timer (or the *least* if they are all *odd*). The best *Right* option is one with *least Left* toll, and, among those, make the same choice of timer (largest even, else least odd).

To find the tallies for a component, suppose that the best Left option has Right tally x_a, and the best Right option has Left tally y_b.

If $x > y$, the game is *hot*, and the tallies are $x_{a+1}y_{b+1}$.

If $x < y$, the game has become *cold*, with tallies $z_0 z_0$, where z is the *simplest* number (with earliest birthday) in the interval

$x < z < y$ if a and b are both even,
$x \leqslant z \leqslant y$ if a and b are both odd,
$x < z \leqslant y$ if a is even and b is odd,
$x \leqslant z < y$ if a is odd and b is even.

even timers *exclude* their tolls,
odd timers *admit* their tolls
as candidates for
the simplest number.

If $x = y$, the game is *tepid*. *Try* $x_{a+1}y_{b+1}$: this is right if $a + 1$, $b + 1$ are both odd, but if they are both even, replace them by 0 (the game has become cold). If just one of $a + 1$, $b + 1$ is even, increase the other (if necessary) by just enough to make it a larger odd number.

There is no room to prove these rules, but we illustrate them with the game *One for Left, Two for Right, Free for All*, a modification of **·007** (*Treblecross*; knock down three contiguous skittles from a row): made *partizan* (and more tidy) by allowing Left to knock down rows with just one skittle, and Right rows of two: and made *selective* by allowing players to attack as many rows as they like in a single move.

In contrast to Treblecross, we can give a complete analysis of this game though you are forgiven if you do not immediately see the pattern in table 16.

Table 16
Tallies for One for Left, Two for Right, Free for All

k	Row of		
	$3k + 1$	$3k + 2$	$3k + 3$
0	1	-1	$0_1 0_1$
1	$1_1 1_1$	$2_1 -1_1$	$0_1 0_1$
2	$1_2 -2_1$	$2_2 -1_2$	0
3	$1_2 1_3$	$2_2 -1_2$	$0_3 0_3$
4	$1_1 1_3$	$2_4 -1_1$	$0_3 0_3$
5	$1_2 1_3$	$2_4 -1_4$	$0_3 0_2$
6	$1_4 1_5$	$2_4 -1_4$	$0_5 0_3$
7	$1_3 1_5$	$2_4 -1_4$	$0_5 0_5$
8	$1_3 1_5$	$2_4 -1_6$	$0_5 0_4$
9	$1_4 1_5$	$2_4 -1_6$	$0_7 0_4$
10	$1_5 1_5$	$2_4 -1_6$	$0_7 0_5$
11	$1_5 1_5$	$2_4 -1_8$	$0_7 0_4$
12	$1_5 1_5$	$2_6 -1_8$	$0_9 0_4$
13	$1_5 1_5$	$2_6 -1_8$	$0_9 0_6$

We calculate the tallies for a row of 16 skittles, which a legal move may reduce of

$1_v 12$ $2_v 11$ $3_v 10$ $4_v 9$ $5_v 8$ $6_v 7$

tallies (earlier entries in table 16)

$1_v 2_2 -1_2$ $0_1 0_{1v} 1_2 1_3$ $1_1 1_{1v} 0$ $2_1 -1_{1v} 2_2 -1_2$ $0_1 0_{1v} 1_2 -2_1$

i.e.,

$$\begin{array}{ccccccc} 1_1 1_3 & 1_3 1_3 & 1_2 - 2_2 & 1_2 1_3 & 1_1 1_1 & 4_2 - 2_2 & 1_2 - 2_1 \\ 1_3 & 1_3 & & 1_3 & 1_1 & & \end{array}$$

are the tallies with best Right toll from Left's point of view. They have no even timer: the least odd is 1. The best Left tolls for Right are

$$\begin{array}{ccccccc} 1_1 & 1_3 & 1_2 & 1_2 & 1_1 & & 1_2 \end{array}$$

and the greatest even timer is 2. Our rules show that $\{..1_1 \mid 1_2..\}$ is tepid and tell us to try $1_2 1_3$. This is correct: although there is an even timer, the other is odd and greater.

The general pattern for the tolls is easy to see: the only exceptions are the Left toll for two skittles and the Right toll for seven. The pattern for the timers is complicated, and further obscured by 22 exceptions. Timers increase at three different rates, one twice another, the third only as the logarithm (to base 2) of the first. If l_i, r_i are the Left and Right timers for a row of $3k + i$ skittles, $i = 1, 2, 3$, then

$$\begin{array}{llll} l_1 = 2\lfloor (k+3)/6 \rfloor + 1\,, & k \geqslant 10\,, & r_1 = 2\lfloor \log_2(k+2) \rfloor - 1\,, & k \geqslant 3\,, \\ l_2 = 2\lfloor \log_2(k+4) \rfloor - 2\,, & k \geqslant 2, & r_2 = 2\lfloor (k+1)/3 \rfloor\,, & k \geqslant 5\,, \\ l_3 = 2\lfloor (k+3)/3 \rfloor - 1\,, & k \geqslant 3\,, & r_3 = 2\lfloor (k+5)/6 \rfloor\,, & k \geqslant 11\,. \end{array}$$

We do not have a misère theory for tardy unions, but urgent unions can be dealt with, both for normal and for misère play, by introducing two new kinds of move: an *overriding* move which *wins* immediately; and a *suiciding* one which *loses* immediately; and we allow infinite tolls for such options. For precise details see WW, (pp. 292–306).

27. Conclusion

We have only scratched the surface of just a few of many aspects of combinatorial games. There is much more for the student to learn, and a great deal for the researcher yet to discover: a few of the many unsolved problems are listed in the Appendix.

Appendix: Unsolved problems in combinatorial games

1. Subtraction games are known to be periodic. Investigate the relationship between the subtraction set and the length and structure of the period. (*Subtraction games* are played with heaps of beans. A move is to take a number of beans from a heap, provided that number is a member of the *subtraction-set*. See Section 16, or WW, pp. 83–86 and 487–498.)
2. Are all finite octal games ultimately periodic? Resolve any number of

outstanding particular cases, e.g., **·6**, **·14**, **·36**, **·64**, **·74**, **·76**, **·004**, **·005**, **·006**, **·007** (One-dimensional tic-tac-toe, Treblecross), **·016**, **·104**, **·106**, **·114**, **·135**, **·136**, **·142**, **·143**, **·146**, **·162**, **·163**, **·172**, **·324**, **·336**, **·342**, **·362**, **·371**, **·374**, **·404**, **·414**, **·416**, **·444**, **·454**, **·564**, **·604**, **·606**, **·644**, **·744**, **·764**, **·774**, **·776**, and Grundy's Game (split a heap into two unequal heaps). Find a game with a period longer than 149459. Explain the structure of the periods of games known to be periodic. If the binary expansion of the kth code digit in the game with code $\boldsymbol{d_0 \cdot d_1 d_2 d_3} \cdots$ is $\boldsymbol{d_k} = 2^{a_k} + 2^{b_k} + 2^{c_k} + \cdots$, where $0 \leq a_k < b_k < c_k < \cdots$, then it is legal to remove k chips from a heap, provided that the rest of the heap is left in exactly a_k or b_k or c_k or . . . non-empty heaps. (See WW, pp. 81–115 and Gangolli and Plambeck 1989.)

3. Examine some hexadecimal games. Obtain conditions for arithmetic periodicity. (*Hexadecimal games* are those with code digits $\boldsymbol{d_k}$ in the interval from **0** to **15**. See WW, pp. 115–116.)
4. Extend the analysis of Domineering to larger boards. For a modest beginning, find the values of 4 × 4 and 4 × 5 boards. See Berlekamp (1988), (WW, pp. 495–498), and section 9. (Left and Right take turns to place dominoes on a checker-board. Left orients his dominoes North–South and Right East–West. Each domino exactly covers two squares of the board and no two dominoes overlap. A player unable to play loses.)
5. Analyze positions in the game of Go (compare Berlekamp 1988).
6. Is Go-Moku (Five-in-a-Row, Go-Bang, Pegotty) a win for the first player?
7. Complete the analysis of impartial Eatcakes (WW, pp. 269, 271, 276–277). (Played with a number of rectangles, $m_i \times n_i$; a move is to remove a strip $1 \times n_i$ or $m_i \times 1$ from each rectangle, either splitting it into two rectangles, or reducing the length or breadth by one. Winner removes the last strip.)
8. Complete the analysis of Hotcakes (WW, pp. 279–282). (Also played with integer-sided rectangles. Left cuts as many rectangles vertically along an integer line as she wishes, and then rotates one of each pair of resulting rectangles through a right angle. Right cuts as many rectangles as he wishes, horizontally into pairs of integer-sided rectangles and rotates one rectangle from each pair through a right angle.)
9. Develop a misère theory for unions of partizan games. (In a union of two or more games, you move in as many component games as you wish. In misère play, the last player loses.)
10. Extend the analysis of Squares Off (WW, p. 299). (Played with heaps of beans. Move is to take a perfect square (>1) number of beans from any number of heaps. Heaps of 0, 1, 2 or 3 cannot be further reduced. A move leaving a heap of 0 is an overriding win for the player making it. A move leaving 1 is an overriding win for Right, and one leaving 2 is an overriding win for Left. A move leaving 3 does not end the game unless all other heaps are of size 3, in which case the last player wins.)
11. Extend the analysis of Top Entails (WW, pp. 376–377). (Played with stacks of coins. Either split a stack into two smaller ones, or remove the top coin

from a stack. In the latter case your opponent's move must use the same stack. Last player wins. Do not leave a stack of 1 on the board, since your opponent must take it and win!).

12. Extend the analysis of All Square (WW, p., 385). (Played with heaps of beans. A move splits a heap into two smaller ones. If both heap sizes are perfect squares, the player must move again: if he cannot he loses!)
13. Extend the misère analysis of various octal games, e.g., Officers, Dawson's Chess, . . . , and of Grundy's Game see Allemang (1984) and WW, (pp. 411–421). William L. Sibert has completed the analysis of misère Kayles, see Sibert and Conway (1992).
14. Moebius, when played on 18 coins, has a remarkable pattern. Is there any trace of pattern for larger numbers of coins? Can any estimate be made for the rate of growth of the nim-values? (See section 20, and WW, pp. 432–435. Played with a row of coins. A move turns 1, 2, 3, 4 or 5 coins, of which the rightmost must go from heads to tails. Winner makes all coins tails.)
15. Mogul has an even more striking pattern when played on 24 coins, which has some echoes when played on 40, 56, or 64 coins. Thereafter, is there complete chaos? (See references for problem 14. A move turns 1, 2, . . . , 7 coins.)
16. Find an analysis of Antonim with four or more coins (WW, pp. 459–462). (Played with coins on a strip of squares. A move moves a coin from one square to a smaller-numbered square. Only one coin to a square, except that square zero can have any number of coins.)
17. Extend the analysis of Kotzig's Nim (WW, pp. 481–483). Is the game eventually periodic in the length of the circle for every finite move set? Analyse the misère version of Kotzig's Nim. (Players alternately place coins on a circular strip, at most one coin on a square. Each coin must be placed m squares clockwise from the previously placed coin, provided m is in the given *move set*. Complete analysis is only known for a few small move sets.
18. Obtain asymptotic estimates for the proportions of $\mathcal{N}$-, $\mathcal{O}$- and $\mathcal{P}$-positions in Epstein's Put-or-Take-a-Square game (WW, pp. 484–486). (Played with one heap of beans. At each turn there are just two options, to add or take away the largest perfect square number of beans that there is in the heap.)
19. Simon Norton's game of Tribulations is similar to Epstein's game but squares are replaced by triangular numbers. Norton conjectures that there are no $\mathcal{O}$-positions, and that the $\mathcal{N}$-positions outnumber the $\mathcal{P}$-positions in golden ratio. This is true up to 5000 beans.
20. Complete the analysis of D.U.D.E.N.E.Y. (Played with a single heap of beans. Either player may take any number of beans from 1 to Y, except that the immediately previous move must not be repeated. When you cannot move you lose. Analysis is easy for Y even, and is known (WW, pp. 487–489) for 53/64 of the odd values of Y.)
21. Schuhstrings is the same as D.U.D.E.N.E.Y., except that deduction of zero is also allowed, but cannot be immediately repeated (WW, pp. 489–490).

2153

sis of The Princess and the Roses (WW, pp. 490–494). ps of beans. Take one bean, or two beans, one from each of heaps.)

nway's and Paterson's game of Sprouts with seven or more spots, isère form with five or more spots (WW, pp. 564–568). (A move wo spots, or a spot to itself by a curve which does not meet any other t or previously drawn curve. When a curve is drawn, a new spot must be placed on it. The valence of any spot must not exceed three.)

24. Extend the analysis of Sylver Coinage (WW, pp. 575–597). (Players alternately name different positive integers, but may not name a number which is the sum of previously named ones, with repetitions allowed. Whoever names 1 loses.)
25. Extend the analysis of Chomp (WW, pp. 598–599). (Players alternately name divisors of N, which may not be multiples of previously named numbers. Whoever names 1 loses.)
26. Extend Uléhla's or Berlekamp's analysis of von Neumann's game from diforests to directed acyclic graphs (WW, pp. 570–572, Uléhla 1980).

Note added in proof

The subject of combinatorial games is a young one, and rapid advances are being made. Since this chapter was first drafted, an A.M.S. Short Course was held in Columbus OH in August 1990, and an M.S.R.I. Workshop in Berkeley CA in July 1994. Serious students of the subject should consult

Combinatorial Games, *Proc. Symp. Appl. Math.* **43** (1991), Amer. Math. Soc., Providence R.I.

and the Proceedings of the Workshop, also to be published by the A.M.S. They should also know that Aviesri S. Fraenkel maintains an up-to-date bibliography which is obtainable from him at The Weizmann Institute of Science, Rehovot 76100, Israel.

We list some recent advances. David Wolfe has found the values of 4 × 4 and 4 × 5 Domineering boards. He and Berlekamp have made significant progress with the analysis of Go endgames: see *Mathematical Go: Chilling Gets the Last Point*, A.K. Peters, 1994. Allis, van den Herik and Huntjens have shown that Go-Moku is a win for the first player. Julian West found loony positions of 2403 coins, 2505 coins, and 33 243 coins in the game of Top Entails. Thane Plambeck has applied Sibert's method to obtain the misère analysis of some more octal games. Fraenkel, Jaffray, Kotzig and Sabidussi have a paper on Kotzig's Nim. Marc Wallace and Alan Jaffray have made progress with the game D.U.D.E.N.E.Y. Daniel Sleator has pushed the normal analysis of Sprouts to 10 spots and the misère analysis to 8. For references, see Fraenkel's Bibliography.

References

Allemang, D.T.

[1984] *Machine computation with finite games,* M.Sc. Thesis (Cambridge University, Cambridge).

Austin, R.B.

[1976] *Impartial and partisan games,* M.Sc. Thesis (University of Calgary).

Berlekamp, E.R.

[1972] Some recent results on the combinatorial game called Welter's Nim, in: *Proc. 6th Conf. on Information Science and Systems, Princeton,* pp. 203–204.

[1974] The Hackenbush number system for compression of numerical data, *Inform. and Control* **26**, 134–140. MR 50#6622.

[1988] Blockbusting and Domineering, *J. Combin. Theory A* 49(1), 67–116.

Berlekamp, E.R., J.H. Conway and R.K. Guy

[1982] *Winning Ways for your Mathematical Plays* (Academic Press).

Bouton, C.L.

[1901–02] Nim, a game with a complete mathematical theory, *Ann. of Math. Princeton (2)* **3**, 35–39.

Conway, J.H.

[1976] *On Numbers and Games* (Academic Press).

Conway, J.H., and H.S.M. Coxeter

[1973] Triangulated polygons and frieze patterns, *Math. Gaz.* **57**, 87–94, 175–183. MR 57#1254–5.

Conway, J.H., and N.J.A. Sloane

[1986] Lexicographic codes: error-correcting codes from game theory, *IEEE Trans. Inform. Theory* **IT-32**(3), 337–348.

Curtis, R.T.

[1976] A new combinatorial approach to M_{24}, *Math. Proc. Cambridge Philos. Soc.* **79**, 25–42.

[1977] The maximal subgroups of M_{24}, *Proc. Cambridge Philos. Soc.* **81**, 185–192.

[1984] The Steiner system $S(5, 6, 12)$, the Mathieu group M_{12} and the "kitten", in: *Computational Group Theory, Durham, 1982,* ed. M.D. Atkinson (Academic Press, London) pp. 353–358. MR 86a:05011.

Dawson, T.R.

[1934] *Fairy Chess Review* (Dec.) p. 94, problem 1603.

[1935] *Caissa's Wild Roses,* p. 13.

Descartes, Blanche

[1953] Why are series musical? *Eureka* **16**, 18–20. Reprinted: 1964, *Eureka* **27**.

Dudeney, H.E.

[1907] *The Canterbury Puzzles and other Curious Problems* (Nelson, London). Reprinted: 1958 (Dover, New York) pp. 118–220.

Epstein, R.A.

[1967] *Theory of Gambling and Statistical Logic* (Academic Press, New York).

Ferguson, T.S.

[1974] On sums of graph games with the last player losing, *Int. J. Game Theory* **3**, 159–167. MR 52#5046.

Fraenkel, A.S.

[1980] From Nim to Go, in: *Proc. Symp. on Combin. Math. and Optimal Design, Fort Collins, CO, 1978,* ed. J. Srivastava, *Ann. Discrete Math.* **6**, 137–156. MR 82e:90117.

Gangolli, A., and T. Plambeck

[1989] A note on periodicity in some octal games, *Int. J. Game Theory* **18**, 311–320. MR 91a:90183.

Grundy, P.M.

[1939] Mathematics and games, *Eureka* **2**, 6–8. Reprinted: 1964, *Eureka* **27**, 9–11.

Grundy, P.M., and C.A.B. Smith

[1956] Disjunctive games with the last player losing, *Proc. Cambridge Philos. Soc.* **52**, 527–533. MR 18, 546.

Guy, R.K.

[1983] Graphs and games, in: *Selected Topics in Graph Theory,* Vol. 2, eds. L.W. Beineke and R.J. Wilson (Academic Press, London) pp. 269–295. MR 87h:90272.

[1986] John Isbell's game of Beanstalk, and John Conway's game of Beans-Don't-Talk, *Math. Mag.* 59, 259–269.

Guy, R.K., and C.A.B. Smith

[1956] The *G*-values for various games, *Proc. Cambridge Philos. Soc.* **52**, 514–526. MR 18, 546.

Hanner, O.

[1959] Mean play of sums of positional games, *Pacific J. Math.* **9**, 81–99. MR 21#3277.

Kenyon, J.C.

[1967a] *A Nim-like Game with Period 349*, Math. Res. Paper 13 (University of Calgary).

[1967b] *Nim-like games and the Sprague–Grundy Theory*, M.Sc. thesis (University of Calgary).

Lenstra, H.W.

[1977] *Nim multiplication, Séminaire de Théorie des Nombres*, exposé No. 11 (Université de Bordeaux).

Loyd, S.

[1914] *Cyclopedia of Tricks and Puzzles* (Morningside Press, New York) p. 232.

Milnor, J.

[1953] Sums of positional games, in: *Contributions to the Theory of Games*, eds. H.W. Kuhn and A.W. Tucker, *Ann. Math. Stud.* **28**, 291–301.

Sibert, W.L., and J.H. Conway

[1992] An analysis of Misère Kayles, *Int. J. Game Theory*, submitted.

Smith, C.A.B.

[1966] Graphs and composite games, *J. Combin. Theory* **1**, 51–81. MR 33#2572.

Sprague, R.P.

[1935–36] Über mathematische Kampfspiele, *Tôhoku Math. J.* **41**, 438–444. Zb 13, 290.

[1947] Bemerkungen über ein spezielle Abelsche Gruppe, *Math. Z.* **51**, 82–84. MR 9, 330–331.

Steinhaus, H.

[1925] Definicje potrzebne do teorji gry i pościgu, *Myśl. Akad. Lwów* **1**(1), 13–14. Reprinted: 1960, Definitions for a theory of games and pursuit, *Naval Res. Logist. Quart.* 7, 105–108.

Uléhla, J.

[1980] A complete analysis of von Neumann's Hackendot, *Int. J. Game Theory* **9**, 107–115.

von Neumann, J.

[1923] Zur Einführung der transfiniten Zahlen, *Acta Litt. Acad. Ser. Szeged X.* **1**, 199–208.

Welter, C.P.

[1952] The advancing operation in a special abelian group, *Nederl. Akad. Wetensch. Proc. Ser. A* **55** [= *Indag. Math.* **14**], 304–314. MR 14, 132.

[1954] The theory of a class of games on a sequence of squares, in terms of the advancing operation in a special group, *Nederl. Akad. Wetensch. Proc. Ser. A* **57** [= *Indag. Math.* **16**], 194–200. MR 15, 682; 17, 1436.

CHAPTER 44

The History of Combinatorics

Norman L. BIGGS

Department of Statistical and Mathematical Sciences, The London School of Economics and Political Science, University of London, Houghton Street, London WC2A 2AE, UK

E. Keith LLOYD

Faculty of Mathematical Studies, University of Southampton, Southampton SO17 1BJ, UK

Robin J. WILSON

Faculty of Mathematics and Computing, The Open University, Walton Hall,
Milton Keynes MK7 6AA, UK

Contents

HANDBOOK OF COMBINATORICS
Edited by R. Graham, M. Grötschel and L. Lovász

Introduction

The history and development of combinatorics cannot be covered completely in a single chapter, partly because the explosive growth of the subject in the last forty years makes it impossible to give a definitive account of recent developments. Fortunately, the other chapters of this Handbook contain many historical remarks and from them a fairly clear idea of the main events in the recent history of the subject can be pieced together. Our aim here is to survey the field, tracing the development of major themes from the earliest times, and showing how current research has evolved from older problems. Inevitably some topics have been partly or completely overlooked, and some mathematicians have been slighted by the omission of their contributions. Nevertheless, it is hoped that an overview of the entire subject, from a historical point of view, will add some new insights to the story so comprehensively described in the rest of this Handbook.

1. Combinatorics in antiquity

It is strange that there is almost no material relevant to combinatorics in the literature of the classical Western civilizations. All the evidence points to the fact that the originators of the subject came from the East. The Chinese have a minor claim, through their interest in magic squares, but the main stimulus came from the Hindus.

The study of ancient Hindu texts is a difficult subject. In many cases it is impossible to assign firm dates, or to separate the original text from later commentaries and embellishments, and consequently some modern Indian historians have made exaggerated claims for the priority of the Hindus in developing various branches of higher mathematics. However, it does seem clear that the basic ideas of choosing and arranging were so intimately related to Hindu culture that a gradual mathematical development of these topics was inevitable. For example, the formula

$$n(n-1)(n-2)\cdots 3\cdot 2\cdot 1$$

for the *number of permutations of an n-set*, and the formula

$$\frac{n(n-1)\cdots(n-k+1)}{k(k-1)\cdots 2\cdot 1}$$

for the *number of k-subsets of an n-set*, were known to Bhaskara around 1150 and probably to earlier mathematicians such as Brahmagupta (sixth century). Special cases of these formulae may be found in texts dating back to the second century BC; further details are given by Biggs (1979).

The *magic square* of order three, see fig. 1, may be reliably traced to Chinese writings of the first century AD, but claims that it was known in 2200 BC are unjustified. Its compelling fascination in times when even the simplest arithmetic

4	9	2
3	5	7
8	1	6

Figure 1.

was a matter for wonderment can easily be imagined. There is no evidence that the Chinese made any further progress in the general study of magic squares until the period 900 to 1300 AD. During this time, the scholars of both China and Islam made extensive studies of magic squares, and were able to construct squares of any given order by a variety of methods. During the 13th century in particular, there was exchange of knowledge between these two cultures; this is exemplified by the discovery in 1956, on the outskirts of the Chinese city of Xi'an, of some iron tablets bearing 6 × 6 magic squares inscribed in East Arabic numerals [see Li and Du (1987) for an illustration of one of them]. The mystical overtones persisted, and survived the transmission of this knowledge to the West, through the Byzantine Greek mathematician Moschopolous, about 1315.

Another fascinating combinatorial object which seems to have filtered westwards around this time is the triangle of *binomial numbers*, see fig. 2. These numbers occur naturally in two contexts: as the number of subsets of a set (known to the Hindus, as already mentioned), and as the binomial coefficients which occur in the Hindu method for the extraction of roots. Thus the triangle itself may well have been known to the Hindus, although the earliest definite instances occur in 13th century works. Hughes (1989) has pointed out that Jordanus de Nemore (fl. AD 1225) discussed its construction and use in his *De Arithmetica* (Book IX, Proposition 70). It also occurs in Arabic works, such as that of al-Tusi in 1265 (Ahmedev and Rosenfeld 1963). The same arrangement appears in Chinese texts around 1300, some of which indicate that it derives from writings (now lost) of Jiă Xiàn (= Chia Hsien = 賈憲) circa 1100 (see Li and Du 1987).

Pascal's famous treatise (1665) on what has come to be known as *Pascal's triangle* was thus by no means the earliest work on the subject; a detailed history of the triangle is given by Edwards (1987). Pascal's treatise is distinguished by the fact that it gives a "modern" deductive treatment of a range of topics related to the binomial numbers, and uses a form of the *principle of mathematical induction*. One notable stimulus for Pascal's work was the problem of predicting results of games of chance. Indeed, such problems provide an important link between the scholars of mediaeval times and modern mathematicians, and this link is

```
        1
      1   1
    1   2   1
  1   3   3   1
1   4   6   4   1
       . . .
```

Figure 2.

especially relevant in the field of combinatorics. An excellent account may be found in a book by David (1962).

2. Origins of modern combinatorics

Pascal's *Traité* (1665) may be said to mark the beginning of combinatorics as we know it today. To a lesser extent Leibniz should also be given some credit for originating several of the main strands of the subject. His *Dissertatio de Arte Combinatoria* (Leibniz 1666), deals only marginally with combinatorics, and is based on work of Lull and others (Knobloch 1979). But we owe to him the suggestion (in a letter to Johann Bernoulli) that it would be rewarding to study the *partitions of integers*, and although he published little on this subject there are many unpublished manuscripts of his which deal with it (Knobloch 1974). This question was later taken up by Euler, who used some of Leibniz's ideas; another letter of Leibniz, written to Huygens in 1679, contains a rather vague reference to a "geometry of position". When Euler solved the problem of the Königsberg bridges (section 5), his friend Ehler (see Sachs et al. 1988) pointed out to him that his work was relevant to the ideas of Leibniz, and Euler subsequently mentioned this in his paper (Euler 1736). In 1833, Gauss referred to the geometry of position as a neglected subject, to which only Euler and Vandermonde had given a "feeble glance" (Gauss 1867, p. 605).

It was during the 17th century that advances in algebraic notation led to a clearer understanding of a fundamental link between algebra and combinatorics: the observation that the expansion and collection of terms in a product of algebraic expressions corresponds to the listing of combinatorial objects of a certain kind. (For example, as already noted in section 1, the binomial expansion can be interpreted as a rule for finding the number of ways of choosing k objects from a set of size n.) This idea was known to Pascal and Leibniz, and a version of it has been ascribed to the English mathematician Harriot, around 1600. De Moivre (1697) carried the idea a step further when he proved the *multinomial theorem*, giving the rule for finding the coefficients in the expansion of

$$(x_1 + x_2 + \cdots + x_r)^n .$$

Another of his discoveries was a form of the *principle of inclusion and exclusion*, which he used (de Moivre 1718) to derive the formula

$$D_n = n! \sum_{r=0}^{n} \frac{(-1)^r}{r!}$$

for the number D_n of *derangements* of n objects. This result had been obtained previously in other ways by Nikolaus Bernoulli and Montmort; see Takács (1981) for a detailed discussion. An account of the life and work of de Moivre is given by Schneider (1968–9).

Of course, the contributions of Euler overshadow everything else in the 18th century. His seminal ideas in graph theory will be noted in section 5, but he also worked in the areas of combinatorics related to *partitions* and *latin squares*. In the first of these areas he made remarkable progress, using the algebraic technique mentioned in the previous paragraph. In his book (Euler 1748) he considered the product

$$(1+x^{\alpha}z)(1+x^{\beta}z)(1+x^{\gamma}z)(1+x^{\delta}z)\cdots,$$

where $A=\{\alpha,\beta,\gamma,\delta,\ldots\}$ is a set of distinct positive integers. Each choice of m integers from the set A summing to n is a partition of n into m distinct parts, each of which is in A. Hence the coefficient of x^nz^m in the product is just the number of such partitions. Similarly, if α may occur more than once as a part, the factor $(1+x^{\alpha}z)$ is replaced by $(1-x^{\alpha}z)^{-1}$, since

$$(1-x^{\alpha}z)^{-1}=1+x^{\alpha}z+(x^{\alpha})^2z^2+(x^{\alpha})^3z^3+\cdots,$$

and the term $(x^{\alpha})^rz^r$ corresponds to r occurrences of α in the partition. In particular, this leads to the formula now known as the *partition-generating function*:

$$\sum_{n=0}^{\infty}p(n)x^n=\prod_{n=1}^{\infty}(1-x^n)^{-1},$$

where $p(n)$ is the total number of partitions of n (see section 4). In this way Euler obtained many interesting identities concerning infinite products and infinite series; another of his examples is given in section 4. He also studied the relationship between partitions and *symmetric functions*.

Euler's interest in latin squares was more transient. In a famous paper (Euler 1782), he posed the following problem:

> "If there are 36 officers, one of each of six ranks from each of six different regiments, can they be arranged in a square in such a way that each row and column contains exactly one officer of each rank and one from each regiment?"

In modern terminology the conjecture is concerned with the existence.of a pair of *orthogonal latin squares* of order 6 (see section 6). Euler was unable to find a solution and he conjectured that no solution exists, not only in the case $n=6$ but generally when $n\equiv 2 \pmod 4$. A solution for $n=4$ corresponds to a well-known arrangement of the sixteen "court" cards in a standard pack of playing cards, and had been published much earlier, in, for example, Ozanam's *Mathematical Recreations* (1725). Euler generalized this arrangement and constructed solutions for many other values of n. (See section 6 for later progress on this topic.)

The formal methods which Euler developed in the study of partitions were developed by Hindenburg (1796) and his collaborators. They used a notation for dealing with symmetric functions and related topics which was so complicated that their work has not been much studied by later scholars. Around the same time, practical mathematicians began to use combinatorial ideas in everyday problems. For example, Peter Nicholson (architect, carpenter, builder and "private teacher of the mathematics") published 250 pages of *Essays on the Combinatorial Analysis* (1818), and this is probably the first book in English on the subject.

A more profound influence on the mathematical development of combinatorics was the study of permutations and their algebraic properties. In describing the properties of what are now called *groups of permutations*, Lagrange, Galois and Cauchy opened up the way for the eventual integration of the subject into the mainstream of modern mathematics. For example, Cauchy (1815) was probably the first to prove formally that exactly half of the permutations of n objects can be expressed as the product of an even number of transpositions. But special cases of this result had been known to English church bellringers for well over a century before that, because the curious rules governing the ringing of changes inevitably lead to a study of transpositions (White 1983).

3. Formal methods of enumeration

Problems of enumeration go back to antiquity, but many were solved by ad hoc methods (if they were solved at all), and it was not until about the end of the 17th century that systematic methods of solution began to be developed. In his work on the multinomial theorem (see section 2), de Moivre (1697) also discussed the reversion (compositional inversion) of series (see below). Later (de Moivre 1718), he used generating functions to solve what are now called *homogeneous difference equations with constant coefficients*, i.e., relations of the form

$$a_n = \alpha_1 a_{n-1} + \alpha_2 a_{n-2} + \cdots + \alpha_r a_{n-r} ,$$

where the α_i are constants. The term *recurrence* is also due to de Moivre (1722), and to him must be ascribed the first general study of the subject. Of course, special cases had been studied before, for example by Leonardo of Pisa (Fibonacci) in his *Liber Abaci* (1202). Fibonacci's problem concerned the breeding of pairs of rabbits, and this led him to a sequence in which each successive term is the sum of the two previous terms. He remarked that this rule enables the sequence to be continued indefinitely but, not surprisingly, he did not give an expression for the general term in the sequence. The terms in the sequence are now called *Fibonacci numbers*. If the nth term is denoted by f_n, the rules for forming the terms are (in modern notation)

$$f_1 = 1 , \qquad f_2 = 1 , \qquad f_n = f_{n-1} + f_{n-2} , \qquad n \geq 3 .$$

De Moivre (1730) gave the explicit solution

$$f_n = \{(\tfrac{1}{2} + \tfrac{1}{2}\sqrt{5})^n - (\tfrac{1}{2} - \tfrac{1}{2}\sqrt{5})^n\}/\sqrt{5} .$$

The use of generating functions in enumeration requires the manipulation of power series, and many results in the 18th century were obtained by purely formal methods. It was taken for granted that a formal power series defines a function, and the series were manipulated without regard to questions of convergence.

Some of the techniques are still remembered today, but others have now been largely forgotten, being swamped in the wake of classical analysis as it gathered momentum in the 19th century. For example, a problem of great importance in manipulating series is that of *reverting power series*, that is, given the equation

$$z = t + V_2t^2 + V_3t^3 + V_4t^4 + \cdots ,$$

it is required to express t in the form

$$t = z + W_2z^2 + W_3z^3 + W_4z^4 + \cdots .$$

There are several ways to do this; de Moivre (1698) obtained results on this problem and there is also much interesting material in Arbogast's book (1800). The latter's material was used by some later mathematicians, including Cayley and Glaisher, but it is barely remembered today. One method which is still used is the Lagrange inversion formula (Lagrange 1770) which states that if the coefficients $U_{n,m}$ are defined by

$$(1 + V_2t + V_3t^2 + \cdots)^{-n} = U_{n,0} + U_{n,1}t + U_{n,2}t^2 + \cdots ,$$

then the coefficients W_n are related to them by

$$W_n = U_{n,n-1}/n .$$

In a series of papers in the 19th century, Blissard (a country clergyman) used techniques quite alien to the thinking of analysts. These methods included expanding series and then at suitable points replacing powers by subscripts; similar techniques were also used by Lucas (1877). The *Blissard* (or *umbral*) *calculus*, as it was later called, was largely ignored except by a few devotees, including Bell and Riordan, but in recent times Rota and his school have been putting the subject on a more rigorous basis – see, for example, Roman (1984). Bell (1938) gave an account of the method, together with details of (and references to) Blissard's life and work.

The functions now known as *permanents* first appeared in the literature in papers of Binet (1812) and Cauchy (1812); a detailed account of the subject, including its history, is given by Minc (1978, 1983). Many of the early papers involved identities between determinants and permanents, but eventually it was realized that a number of enumeration problems can be stated in terms of evaluating the permanents of various (0, 1)-matrices. For example, if $\mathrm{F} = (A_1, \ldots, A_m)$ is a finite family of subsets of the finite set $S = \{s_1, \ldots, s_n\}$, then the number of transversals of F (see section 7) is equal to the permanent of the incidence matrix $P = (p_{ij})$ of the system, where $p_{ij} = 1$ if $s_i \in A_j$, and $p_{ij} = 0$ otherwise. Unfortunately, the evaluation of permanents is a hard problem (see chapter 29), so some of this work is only of theoretical interest.

Another type of problem which can be expressed in terms of permanents is that in which it is desired to find the number of rearrangements of an ordered set of elements, subject to certain specified restrictions as to which elements may go in which positions. Special cases of such problems may be solved without reference to permanents, and include the *derangement problem* (*problème des rencontres*) mentioned in section 2, and the *problème des ménages*. The history of the latter problem is discussed by Dutka (1986). As mentioned in section 2, one technique for solving such problems is the *principle of inclusion and exclusion*, which is an example of a *sieve method* – a technique in which the elements of a set larger than the one of interest are first listed or counted, and then various subsets are either deleted or added until the set required, or the number of elements in it, is obtained. One such method is used in number theory and employs a *Möbius function* defined on the positive integers. In the mid-1930s the concept of a Möbius function was extended to other posets independently by Weisner and P. Hall and, a few years later, generalized by Ward. The idea was taken up enthusiastically by Rota (1964) – see chapter 21.

In the late 19th and early 20th centuries, major contributions were made to enumeration by MacMahon. There is insufficient room here to describe all his contributions but further details may be found in his books (MacMahon 1915–16) and collected papers (MacMahon 1978/86) as well as in the commentaries provided by Andrews in the latter works. One result which deserves mention, however, is the *master theorem*.

MacMahon's Master Theorem. *Let* $x = (x_1, x_2, \ldots, x_n)^{\mathrm{T}}$ *and* $y = (y_1, y_2, \ldots, y_n)^{\mathrm{T}}$ *be column vectors connected by the matrix equation* $y = Ax$, *where* A *is an* $n \times n$ *matrix. Also let* $\Delta = \det(I - AX)$, *where* X *is the diagonal matrix* $\operatorname{diag}(x_1, x_2, \ldots, x_n)$, *Then the coefficient of*

$$x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n}$$

in the expansion of

$$y_1^{i_1} y_2^{i_2} \cdots y_n^{i_n}$$

is equal to the coefficient of

$$x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n}$$

in the expansion of $1/\Delta$.

MacMahon applied this theorem to various problems such as counting permutations of multisets in which no element remains in its original position; see also Cartier and Foata (1969).

There are many enumeration problems in which the equivalence of structures is defined in terms of the action of a finite group of symmetries. The history of the following *orbit-counting theorem*, sometimes incorrectly known as *Burnside's lemma* or, more appropriately, as the *Cauchy–Frobenius lemma*, is discussed in some detail by Neumann (1979). The theorem lies at the basis of many results on enumeration under group action.

Orbit-Counting Theorem. *If the elements of a finite group G act as permutations of the elements of a finite set D, then the number of orbits under the action is given by*

$$\frac{1}{|G|}\sum_{g\in G}|\mathrm{fix}(g)|\,,$$

where fix(g) *is the subset of elements in D which are fixed by the action of g, and $|X|$ denotes the number of elements in X.*

The theorem was used by Redfield, who, inspired by some of MacMahon's results, did important work in this area. Unfortunately, his remarkable paper (Redfield 1927) remained almost unnoticed until about 1960, but since then it has come to be regarded as an outstanding contribution to the subject. Redfield introduced some polynomials which he called *group reduction functions*, and performed various operations upon them in order to solve a number of counting problems. He stated a theorem, now known as the *Redfield–Read superposition theorem* [since it was discovered independently by Read (1959)], which can be used in a great variety of enumeration problems. In the 1980s, valuable unpublished work by Redfield came to light; see Lloyd (1988) for further information. Biographical details of Redfield are given by Lloyd (1984).

In the late 19th century, both chemists and mathematicians began to study problems of counting chemical isomers (see chapter 38). In particular, Cayley, in his work on trees (see section 5) considered the alkane series C_nH_{2n+2}. The early methods for isomer enumeration were rather cumbersome and many errors appeared in the early work, but in the 1920s mathematical tools were developed which led to significant progress. Had it been read and understood at the time, Redfield's paper (1927) might well have been a major influence in this area, but only in the 1980s did chemists begin to realize the usefulness of his methods. More influential were the works of Lunn and Senior (1929) and Pólya. Pólya's ideas were set out in a number of papers in the mid-1930s, culminating in a lengthy and famous paper (Pólya 1937) translated into English half a century later (Pólya and Read 1987). Pólya combined the use of generating functions with the Orbit-Counting Theorem and he applied his methods to counting various graphs and chemical compounds. Some of his ideas had already appeared in Redfield's paper (in particular, Pólya's *cycle index* is Redfield's group reduction function), and these ideas were further developed by de Bruijn. A detailed discussion of the interrelationships of work in this area is given by Read (1968); see also Read's chapter in Pólya and Read (1987).

The present state of the art of enumeration is described in chapters 21 and 22.

4. Partitions and symmetric functions

After Euler's work on partitions and that of Hindenburg and his colleagues, both mentioned in section 2, there was little further progress until the 1840s, when a number of mathematicians began to look at the subject. In particular, Warburton sought a method for determining the number of partitions of a given number, and he communicated some of his results to De Morgan. The latter presented an account of Warburton's work to the Royal Society in 1847 and, soon afterwards, Warburton himself published a paper (1842–9) on the subject. Although the paper did not contain outstanding results, according to MacMahon (1896–7) it did have the effect of bringing the subject to the attention of mathematicians such as Herschel (1850). There are numerous recurrence relations between the numbers $P_r(n)$ of partitions of n into r parts, and De Morgan, Warburton and Herschel attempted to solve such relations. Herschel expressed $P_r(n)$ in terms of certain functions known as *circulating functions*.

Circulating functions were also used by Cayley and Sylvester, who (unlike Herschel) started with generating functions and sought an expression for the coefficient of the general term. Sylvester's work on partitions spanned many years, during which time he was often sidetracked by other researches, but by making use of Cauchy's work on the theory of residues, he obtained and published an expression for the coefficient of x^n in an arbitrary rational function (Sylvester 1855–7), although it was many years before he published a proof. This work was described by MacMahon (1896–7) as "incomparably the finest contribution that has ever been made to combinatory analysis". Glaisher (1875, 1909) also worked on partitions rather intermittently, but his approach was rather different from earlier writers, and he made use of the methods of Arbogast (1800) for calculating coefficients in series expansions.

In contrast to the analytical methods of Sylvester, elegant proofs were obtained by using diagrams first published by Sylvester, but attributed by him to Ferrers. If the parts λ_r of a partition are arranged in non-increasing order ($\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k > 0$), then the *Ferrers diagram* is obtained by placing λ_r marks, such as dots or ones, in the rth row of the diagram. For example, the diagram for the partition $3+3+2+1$ is shown in fig. 3.

A rather remarkable result on partitions states that

$$(1-x)(1-x^2)(1-x^3)(1-x^4)\cdots=\sum_{n=0}^{\infty}(-1)^n x^{n(3n\pm1)/2}\,.$$

This is equivalent to the statement that except for numbers m of the form $\frac{1}{2}n(3n\pm1)$, the partitions of m into an even number of parts are equinumerous with those into an odd number of parts, and in the exceptional cases the two numbers differ by 1. This result was observed by Euler at an early stage in his work on partitions, but he was unable to prove it until some years later (Euler 1751). Franklin (1881) gave a very neat proof of it using Ferrers diagrams: his

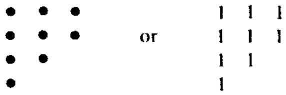


Figure 3.

proof involves moving some of the dots in the diagrams so that an explicit bijection is obtained between certain classes of partitions. Durfee (1882–3) also used Ferrers diagrams: he expressed each diagram as the union of the largest square of dots in the diagram, now known as the *Durfee square*, and two other partitions. This idea enables combinatorial proofs to be given of certain partition-generating function identities.

The number $p(n)$ of partitions of n increases rapidly with n. Values were calculated by Euler to at least $n = 69$, but a longer table, going as far as $p(200) = 3\,972\,999\,029\,388$ was calculated by MacMahon and included in the truly remarkable paper of Hardy and Ramanujan (1918). In that paper, the authors obtained an almost unbelievable result which expresses $p(n)$ as the nearest integer to an expression involving an assortment of square roots, derivatives, exponentials and $(24q)$th roots of unity. Studying MacMahon's table, Ramanujan (1919) was able to prove a number of congruence properties of $p(n)$. For example, he showed that $p(5n + 4) \equiv 0 \pmod 5$ and $p(7n + 5) \equiv 0 \pmod 7$. These are special cases of a more general conjecture which he made but which eventually proved not to be correct. The reader is referred to Andrews (1976) for further details.

The idea of a partition can be extended to higher dimensions and much work was done in this area by MacMahon; see, for example, MacMahon (1915–16). If the dots in a Ferrers diagram are replaced by positive integers subject to the restriction that numbers are non-increasing along each row and also down each column, then the set of integers is a *plane partition* of their sum. For example, the diagram in fig. 4 is a plane partition of 26. Further details are given in chapter 21.

Hammond (1882, 1883) introduced some *differential operators* which act on symmetric functions, and these were enthusiastically employed by MacMahon to process the symmetric functions upon which much of his extensive researches in combinatorics were based (MacMahon 1915–16, 1978/86). More recently, David et al. (1966) used Hammond operators to calculate tables of symmetric functions.

In the first of a series of nine papers written over many years, Young (1901) introduced the idea of a *tableau*. Tableaux correspond to special plane partitions in which the parts are consecutive integers $1, 2, \ldots, m$, for some m, and although Young introduced them in the context of invariant theory, Frobenius (1903) pointed out that Young's methods are closely related to his own work on group representation theory. A problem considered in that subject is to obtain information about the representations of a given group in terms of representations of suitably chosen subgroups. For the symmetric group S_n, the appropriate subgroups are those of the form

$$S_{\lambda_1} \times S_{\lambda_2} \times \cdots \times S_{\lambda_k}\,, \quad \text{where } \sum_r \lambda_r = n\,.$$

5	2	1	1
4	2	1	1
3	2	1	
2	1		

Figure 4.

By combining his ideas with contemporary work of Frobenius and Schur, Young obtained a complete classification of the irreducible representations of S_n over the complex numbers. Young's papers are not easy to read, but in recent years the influence of his work has been felt in a number of areas. His collected papers have been published (Young 1977) and in a review of that work, Andrews (1979) lists 121 papers which cite Young and which were published within the previous fifteen years.

An algorithm of Robinson (1938), rediscovered by Schensted (1961), establishes bijections between certain sets of matrices and certain tableaux. This algorithm enabled Schensted to produce extensions of a famous result of Erdős and Szekeres (see section 7), such as the following.

Theorem. *The number of permutations of* $1, 2, \ldots, m$, *with longest increasing subsequence of length c and longest decreasing subsequence of length r, is equal to* $\sum_\mu f_\mu^2$, *where* f_μ *is the number of standard Young tableaux of shape* μ, *and the summation is over all partitions* μ *of m with c parts and largest part r.*

The *invariant theory of binary forms* is a subject which flourished in the 19th century and, in particular, Sylvester tried to link it with chemistry (see section 5). Since then it has been pronounced dead on many occasions, but it persists in coming back to life and, in the words of Kung and Rota (1984), "the artillery of combinatorics" is now being aimed at the subject. The reader is referred to that paper for further information and references.

5. The development of graph theory

[A fuller account of this topic appears in the book *Graph Theory 1736–1936* (Biggs, Lloyd and Wilson 1976) which includes extracts (translated into English where necessary) from many of the works cited below. Such a work is annotated below in the form (BLW nX), meaning that an extract from it appears as extract X in chapter n of the book.]

The subject of graph theory originated with Euler's solution of the *Königsberg bridge problem*, which asks for a route crossing each of the seven bridges of Königsberg just once. On 26 August 1735, Euler presented a paper to the St. Petersburg Academy of Sciences, proving that the problem is impossible, and showing how his method can be extended to any number and arrangement of islands and bridges. In particular, he formulated necessary and sufficient conditions for the existence of a route crossing every bridge just once, but he seems to have considered it unnecessary to prove the sufficiency in general; the first valid proof of this was given by Hierholzer (BLW 1B).

Euler's paper on the Königsberg problem was written in 1736, and first published in 1741 (BLW 1A), but initially it aroused little interest. Although his methods were essentially graph-theoretical, he did not use graphs as such, and the graph usually used to solve the problem seems not to have appeared until 150 years later (Rouse Ball 1892). Indeed, the problem was not well known until the end of the 19th century, when Lucas (1882) and Rouse Ball (1892) included it in their books on recreational mathematics, although a French translation of Euler's paper had been published earlier by Coupy (1851).

A recreational puzzle related to the Königsberg bridges problem is that of finding the smallest number of pen-strokes needed to draw a given diagram with no line repeated. Poinsot (1810) showed that a single stroke is sufficient for an odd number of points all joined in pairs (the *complete graph* K_n, n odd), but not for an even number of points. Such problems were also discussed by Listing in his *Vorstudien zur Topologie* (1847; BLW 1C), the first publication to use the word *topology*, but the connection between them and the Königsberg problem seems to have remained unnoticed until Lucas's book. The term *Eulerian graph* for a graph which can be drawn with a single pen-stroke is due to König, and appeared in his pioneering book (König 1936). The origins of topology are discussed by Pont (1974).

Another type of traversal problem involves finding a cycle through every vertex of a graph, rather than a route passing along each edge, as above. An example of such a problem is the *knight's tour problem* which asks for a sequence of knight's moves on a chessboard, visiting every square and returning to the starting-point. Although solutions of this problem had been known since the 14th century, the problem was not subjected to mathematical analysis until 400 years later, with papers of Euler (1759) and Vandermonde (BLW 2A). The first general discussion of vertex-traversal problems was given by Kirkman (BLW 2B) who asked which polyhedra allow a cycle passing through each vertex, and described a general class of polyhedra for which no such cycle exists. Graphs which allow such a cycle are now called *Hamiltonian graphs*. Hamilton became fascinated by paths and cycles on a dodecahedron (BLW 2C) as an offshoot of his own work on non-commutative algebra, and of his *icosian calculus* in particular. Such considerations gave rise to a recreational puzzle, the *Icosian Game*, in which the object was to find paths and cycles on a "flat dodecahedron", satisfying certain specified conditions. A version of this game played on a solid dodecahedron was later marketed under the name *A Voyage round the World*.

Unlike the Eulerian problem, where necessary and sufficient conditions for the existence of a trail are easy to find, the general Hamiltonian problem has remained intractable, and the problem of determining whether a given graph is Hamiltonian has been shown to be NP-hard; see chapter 29. However, there are several necessary conditions or sufficient conditions for a graph to be Hamiltonian, such as the sufficient conditions of Dirac (1952), Ore (1960) and others, and a result of Fleischner (1974) that the square of any 2-connected graph is Hamiltonian. Further results on Hamiltonian graphs may be found in a survey by Bermond (1978) and in chapter 1.

Whereas Eulerian and Hamiltonian graphs arose out of recreational puzzles, the study of *trees* emerged from a problem in the differential calculus. In 1857, Cayley (BLW 3A) expressed the problem in terms of *rooted trees* and used generating functions to determine the number of such trees with a given number of edges. Although the concept of a tree had been used implicitly ten years earlier by von Staudt and by Kirchhoff, Cayley was the first to use the term in print.

Cayley's methods for enumerating trees were initially very cumbersome, but Jordan (BLW 3B) simplified the procedures substantially by introducing the concepts of *centroid* and *bicentroid*, and *centre* and *bicentre*, for a given tree. Using them, Cayley (BLW 3C) was able to count trees by starting from their centre or centroid and working outwards. In this way he counted both unrooted trees and various types of chemical molecules, such as the family of alkanes C_nH_{2n+2} (BLW 4B). Closely involved with Cayley's work were Sylvester and Clifford. Both made important contributions to the study of invariants, and Sylvester wrote a note (BLW 4C) and a long paper (Sylvester 1878) aiming to link invariant theory with chemistry by describing an analogy between binary quantics and chemical atoms. To represent this connection diagrammatically, Clifford introduced graphic notation, or graphs for short, and Sylvester's note used the word *graph* (in the graph-theory sense) for the first time; see also chapter 38 for further historical remarks on this topic.

Other tree-counting problems were solved later. In 1889, Cayley announced his n^{n-2} formula for the number of *labelled* trees with n vertices, but verified it only for $n \leqslant 5$. A proof was later given by Prüfer (BLW 3D). The much-needed breakthrough in enumerative techniques occurred in the 1920s and 1930s (see section 3), with the work of Redfield (1927), Lunn and Senior (1929) and Pólya (1937). In particular, Pólya enumerative theory was a milestone for the counting of graphs and chemical molecules. Further information on the enumeration of isomers is given in chapter 38.

One of the most fundamental results in the study of polyhedra, and consequently of graphs embedded in the plane, is *Euler's polyhedral formula* $v - e + f = 2$, relating the numbers of vertices, edges and faces of a polyhedron or connected planar graph. This formula first appeared in a letter on polyhedra written by Euler to Goldbach in November 1750 (BLW 5A), but Euler failed to produce a valid proof of it. The result is sometimes incorrectly attributed to Descartes, who obtained an expression for the sum of the angles of all the faces of a polyhedron. Although Euler's formula can be deduced from this expression, there is no evidence that Descartes did so.

The first correct proof of Euler's formula involved the metrical properties of spherical polygons, and was found by Legendre (1794). A topological (non-metrical) proof was given in 1813 by Cauchy (BLW 5B), who projected the polyhedron onto a plane and used a triangulation method to derive the result for planar maps. Meanwhile, Lhuilier (BLW 5C) showed how Euler's formula leads to a proof that there are only five regular polyhedra. He also investigated the modifications in the formula if the polyhedron has a hole, if its faces are not simply connected, or if the polyhedron is ring-shaped. In this last case, Lhuilier obtained the formula

$$v - e + f = 2 - 2g$$

for a graph embedded on a sphere with g handles.

This last result was the starting-point for an extensive investigation by Listing. In *Der Census räumlicher Complexe* (Listing 1861–2) he studied simplicial complexes, thereby laying the groundwork for Poincaré's development of algebraic topology at the turn of the century. Poincaré showed how complexes can be constructed from basic "cells" – where a 0-cell is simply a vertex, and a 1-cell is an edge. In order to fit the cells together, he adapted a technique of Kirchhoff (BLW 8A), replacing systems of linear equations by the corresponding matrices. These ideas were later developed by Veblen in a series of American Mathematical Society Colloquium Lectures (BLW 8B).

It was already known in the 19th century that certain graphs cannot be embedded in the plane; for example, Möbius's problem of the five princes, and the gas, water and electricity problem (BLW, pp. 115–116 and 142), show that the complete graph K_5 and the complete bipartite graph $K_{3,3}$ are both non-planar. In 1930 Kuratowski proved that these are the "basic" non-planar graphs, in the sense that *every* non-planar graph must contain a subdivision of at least one of them (BLW 8C); a full discussion of the origins of this result is given in Kennedy et al. (1985). This idea has more recently been developed by Glover et al. (1979), who obtained a list of 103 "forbidden subgraphs" for graphs embedded in the projective plane. Furthermore, Robertson and Seymour (1985) have proved that there is a corresponding finite list of forbidden subgraphs for surfaces of *any* genus, although this list may be very large even for surfaces as simple as the torus. Further discussion of work in this area can be found in chapters 5 and 10.

Another aspect of planarity was developed in the 1930s in a series of papers by Whitney. In the first of these (Whitney 1931), he showed that the relationship between a planar graph and its geometrical dual leads to a combinatorial definition of duality which can be used to characterize planar graphs. Extending these ideas led him eventually to the idea of a *matroid*, or *abstract independence structure*, which generalizes ideas of independence in both vector spaces and graphs (Whitney 1935). Indeed, the duality of a matroid is a very natural concept which extends and clarifies the dual of a planar graph. Interest in matroids was slow to develop, but in 1959 Tutte obtained a Kuratowski-type criterion for a matroid to arise from a graph (Tutte 1959). His results paved the way for an explosion of interest in the 1960s and 1970s, fuelled by the discovery of Edmonds and Fulkerson (1965) that the partial transversals of a family of sets give rise to a natural matroid structure. For extensive treatments of matroid theory, see Welsh (1976), Oxley (1992) and chapters 9–11 of this Handbook.

No account of the history of graph theory would be complete without a discussion of the *four-colour problem*. This problem first arose in 1852 when Francis Guthrie noticed that only four colours are needed to colour a map of England and wondered whether this is so for all maps. His brother Frederick approached Augustus De Morgan, who communicated the problem to other mathematicians, including Hamilton. It first appeared in print in 1860 in an unsigned book review by De Morgan (see Biggs 1983).

The four-colour problem was revived in 1878 when Cayley asked at a London Mathematical Society meeting whether it had been solved. In the following year, Kempe produced his celebrated "proof" for the newly-founded *American Journal of Mathematics* (BLW 6B). This "proof" contained some good ideas and was generally accepted until 1890 when Heawood found an error in it. He salvaged enough to prove a *five-colour theorem*, gave a formula for the number of colours needed for maps on a sphere with g handles, and justified his formula for maps on a torus (BLW 6D, 7A). Unfortunately, although this formula gives the number of colours *sufficient* for colouring a map on a surface, his proof that this number is *necessary* for some maps was deficient. Filling the gap proved a difficult task, involving twelve separate cases, but it was eventually completed in 1968. For a full account of this work see Ringel (1974).

Meanwhile, progress on the four-colour problem was also slow and painful. Birkhoff (1913) showed that certain configurations in a map are *reducible*, in the sense that a four-colouring of the rest of the map can be extended to a colouring of the whole map; this idea of reducibility turned out to be crucial in the eventual proof of the theorem. Franklin (1922) used reducibility to prove the theorem for maps with at most 25 countries, and this number was increased over forty years to 95. Finally Appel, Haken and Koch, using ideas of Heesch, proved the *four-colour theorem* in 1976. Although the main idea of the proof can be traced back to Kempe, the details were very complicated, involving the analysis of almost 2000 configurations and the use of hundreds of hours of computing time. An account of their search for a proof is given in Appel and Haken (1977); see also Appel and Haken (1989).

Chromatic graph theory has also developed in other directions: for example, chromatic polynomials were introduced by Birkhoff (1912–3) and critical graphs by Dirac (1952). Other important topics include Brooks's upper bound for the chromatic number of a graph (Brooks 1941), Hadwiger's conjecture (Hadwiger 1943) which has the four-colour theorem as a special case, and two papers of Vizing (1964, 1965) on edge-colourings of graphs. Further information about graph colourings appears in chapter 4.

The origins of *extremal graph theory* can be traced back to a question of Mantel, solved in 1907 by Wythoff (1906–10), and to a paper of Erdős (1938); they determined the maximum number of edges in a graph containing no complete graph K_3 or K_4. The modern development of the subject dates from an important paper of Turán (1941), which solved the corresponding problem for K_n. *Probabilistic graph theory* was also developed in Hungary, in a series of papers by Erdős and Rényi; see, for example, Erdős and Rényi (1960). Their aim was to see how the properties of graphs change as edges are added to a graph at random, and their influence is still strongly felt in the recent development of the subject. For full accounts of random graph theory and extremal graph theory, see chapters 6 and 23.

The above account deals exclusively with finite graphs, but some results can be extended to infinite graphs (see chapter 42) by using set-theoretic results of König (1927) and Rado (1942). If the graph is countable, with finite vertex-degrees, *König's lemma* (in its graphical form) guarantees the existence of a one-way infinite path from any vertex. For uncountable graphs one usually needs a stronger result, known as *Rado's selection theorem*, which is one of the most important tools in infinite combinatorics. A discussion of these results can be found in Thomassen (1983).

6. Configurations and designs

The Chinese studies of magic squares (section 1), and Euler's passing interest in orthogonal latin squares (section 2), may be considered as part of the prehistory of the modern subject of configurations and designs. But the catalyst for the foundation of the theory was geometry, and geometrical ideas pervade the subject to this day.

In 1835, the geometer Plücker remarked that a general plane cubic curve has nine points of inflexion, which lie in threes on twelve lines; furthermore, given any two of the points, one of the twelve lines passes through both of them. In a footnote, he remarked (wrongly) that a system $S(n)$ of n points, arranged in triples in such a way that any two points belong to just one triple, is possible only when $n \equiv 3 \pmod 6$. In 1839 Plücker corrected his error, pointing out that both $n \equiv 1 \pmod 6$ and $n \equiv 3 \pmod 6$ are possible, and he made some remarks about other systems of this kind; these references were brought to light by De Vries (1984). It seems likely that Plücker's comments were noted by Sylvester who communicated them to Woolhouse, the editor of a curious English publication known as the *Lady's and Gentleman's Diary*. What is certain is that Woolhouse proposed the Prize Question for the readers of the *Diary* for 1844 in the following terms:

> "Determine the number of combinations that can be made out of n symbols, p symbols in each; with this limitation, that no combination of q symbols which may appear in any one of them shall be repeated in any other."

In modern terminology, the question asks for the number of blocks in a q-*design* (usually called a t-design in modern terminology, provided repeated blocks are not allowed) with parameters $(n, p, 1)$; Plücker's system $S(n)$ corresponds to the case $p = 3$, $q = 2$. The determination of the number of blocks is simple, but the question of the existence of the design is not. For this reason, Kirkman's paper "On a problem in combinations" (1847), read to the Literary and Philosophical Society of Manchester on 16 December 1846, is truly remarkable, for it showed in effect how to construct the system $S(n)$ whenever $n \equiv 1$ or $3 \pmod 6$. So, the existence problem for $S(n)$ was completely solved.

A little later Kirkman noticed that there is a system $S(15)$ with the property that its 35 triples can be partitioned into seven sets of 5 triples, in such a way that each symbol occurs exactly once in each set of five. Thus was born the famous *fifteen schoolgirls problem*, which apeared as Query VI in the *Lady's and Gentleman's Diary* for 1850:

> "Fifteen young ladies in a school walk out three abreast for seven

> days in succession: it is required to arrange them daily so that no two shall walk twice abreast."

It is sad that Kirkman's name should be remembered primarily for this trifle, because his mathematical papers entitle him to be regarded as the founding father of the theory of designs, rather than as the author of an amusing puzzle. In addition to his solution of the existence problem for $S(n)$, he constructed 2-designs with parameters $(r^2 + r + 1, r + 1, 1)$, now known as *projective planes*, for every prime value of r, and he used cyclic difference sets to construct projective planes with $r = 4$ and $r = 8$. He also found 3-designs with parameters $(2^n, 4, 1)$ and several other special kinds of design. A fuller discussion of Kirkman's life and work is given by Biggs (1981).

Kirkman's work went almost unnoticed at the time. Indeed, some years later the geometer Steiner (1853) revived Plücker's question in an article in Crelle's *Journal*. This is why $S(n)$ is usually known as a *Steiner triple system*, even though the major work on it had been published six years before Steiner's paper.

The remainder of the 19th century saw much work on variants and extensions of the schoolgirls problem. Both Cayley and Sylvester published papers in this area, and Cayley coined the name *tactic* for the general area of configurations and designs. But it was not until the end of the century that the theory of designs was once again the subject of a truly significant paper, the "Tactical memoranda" of Moore (1896). Moore's work is noteworthy for its systematic treatment of the numerical conditions for the existence of designs, and for the use of finite fields to construct various families of designs.

At the end of the 19th century, a new influx of geometrical ideas began to extend and revitalize the subject of "tactic". The idea that geometry can be formulated within a system having only a finite number of points may be traced back at least to Von Staudt (1856–7). The notion was developed by the Italian geometer Fano (1892), who described finite geometries of various dimensions and, in particular, the finite plane with seven points which bears his name. Of course, the *Fano plane* is just the Steiner triple system $S(7)$, and is defined by the axioms for a 2-design with parameters $(7, 3, 1)$, or a projective plane of order 2. The geometrical requirements that two points determine a unique line, and (in plane projective geometry) that two lines meet in just one point, provide the link between geometry and the theory of designs.

This link was the stimulus for American geometers to apply Moore's "Tactical memoranda" to questions of finite geometry. A paper of Veblen and Bussey (1906) continued Fano's work, giving an axiomatic definition of a finite projective geometry in any number of dimensions. They showed that the *Desargues theorem* holds in any such geometry, except possibly in the plane case, and that the *Desarguesian geometries* can be coordinatized by the finite *Galois fields* GF(q). It follows that there are *Desarguesian projective planes* of order q for each prime power q, based on GF(q), but there may also be non-Desarguesian planes. Almost immediately afterwards, Veblen and Wedderburn (1907) showed that certain skewfields constructed by Dickson can be used to coordinatize non-

Desarguesian planes, and they described in some detail a non-Desarguesian plane of order 9.

The existence of finite projective planes is also related to the orthogonal latin squares studied by Euler (see section 2), because a finite projective plane of order n gives rise to a set of $n-1$ mutually orthogonal latin squares (MOLS) of order n. This result was apparently not stated explicitly until it was noticed independently by Bose (1938) and Stevens (1939). Thus Euler's famous conjecture, that there is no pair of orthogonal latin squares of order n whenever $n \equiv 2 \pmod 4$, would imply the weaker result that there are no projective planes with such an order. The Euler conjecture for $n = 6$ was allegedly verified by Clausen around 1842, but his proof was never published, and the first convincing proof was given by Tarry (1900). On the other hand, the work of Moore, Veblen, and their colleagues showed that, when n is a prime power q, there are $q-1$ MOLS of order q. Macneish (1922) remarked that r MOLS of order a and r MOLS of order b can be combined to give r MOLS of order ab; his proof is a reformulation of a construction in Moore's "Tactical memoranda". It follows that, if the prime factorization of n is $p_1^{i_1} p_2^{i_2} \cdots p_r^{i_r}$, then there are at least

$$\min(p_1^{i_1}, p_2^{i_2}, \ldots, p_r^{i_r}) - 1$$

MOLS of order n. Macneish conjectured that this is the maximum number for any n, extending the conjecture of Euler. But Macneish's attempt to prove the Euler conjecture was fallacious, and no further progress was made at that time.

Although the first known recorded use of a latin square in an experimental design was by Cretté de Palluel in 1788 (Street and Street 1988), it was not until the 1920s that interest in latin squares and designs was renewed, as a result of the work of Fisher and Yates on the design of agricultural experiments. They discussed not only the practical applications, but also the theory: for example, Fisher and Yates (1934) completely classified the latin squares of order 6, thereby verifying independently that no orthogonal pair exists. Yates (1936) discussed what are now called *balanced incomplete block designs* (BIBDs), i.e., 2-designs with parameters (v, k, λ) satisfying $v > k$, and Fisher and Yates published a list of known BIBDs in their book (Fisher and Yates 1938). The additional parameters r (number of replications) and b (number of blocks) for a BIBD satisfy the elementary conditions

$$rv = bk\,, \qquad \lambda(v-1) = r(k-1)\,,$$

but it is clear that these conditions are not sufficient for the existence of a BIBD, since, for example, there is no projective plane of order 6 ($v = b = 43$, $k = r = 7$, $\lambda = 1$). Fisher (1940) established the inequality $b \geqslant v$, a surprisingly non-trivial constraint. Around the same time Bose (1939) published a long paper giving an account of everything then known about the construction of BIBDs, and describing several new methods of construction.

Yates (1936) gave a construction which, in present-day terminology, connects a complete set of MOLS with a projective plane. A major step forward was made by Bruck and Ryser (1949), who explicitly considered projective planes as BIBDs with $v = b$ and $\lambda = 1$, and they applied Hasse–Minkowski theory to the incidence matrix of the plane. In this way they proved the non-existence of projective planes of order n for infinitely many values of n, and in particular when $n = 2p$, where p is a prime with $p \equiv 3 \pmod 4$. A little later Chowla and Ryser (1950) applied similar methods to the more general case of symmetric designs ($v = b$, $\lambda \geq 1$). They also discussed the relationships between symmetric designs and earlier studies of Hadamard matrices (Paley 1933, Todd 1933), and cyclic difference sets (Singer 1938, Hall 1947). In general, it is an unsolved problem to find a complete set of necessary and sufficient conditions for the existence of a t-design with parameters (v, k, λ), although Wilson (1975) has shown that the elementary conditions given above are sufficient for a BIBD to exist, provided that v is large enough. In the same vein, Ray-Chaudhuri and Wilson (1971) proved that, for each $n \equiv 3 \pmod 6$, there is a Steiner triple system $S(n)$ which can be partitioned in the manner of Kirkman's schoolgirls problem; similar results had been obtained independently by Lu by 1965 (see Wu et al. 1990).

Towards the end of the 1950s, the problem of constructing a pair of MOLS of order 10 once again became active. It is likely that the availability of electronic computers provided a stimulus for this work, although the first results were obtained without the use of computers. The breakthrough came when Parker (1959a) found two MOLS of order 21, thereby disproving Macneish's extension of the Euler conjecture. Shortly afterwards, Bose and Shrikhande (1959) constructed two MOLS of order 22, and Parker (1959b) found two MOLS of order 10. Finally these three authors combined (Bose et al. 1960) to show that Euler's conjecture is false for *all* $n \equiv 2 \pmod 4$ and $n > 6$. This achievement was front-page-news in the *New York Times* on 25 April 1959.

Another problem which has been studied by computational search techniques is the question of the existence of a projective plane of order 10. The non-existence of a plane of order 6 was proved explicitly by MacInnes (1907); it also follows from Tarry's result on MOLS, and the Bruck–Ryser theorem. The order 10 is not excluded by the Bruck–Ryser theorem, however, and the existence of a pair of MOLS of order 10 is inconclusive. After much effort, it was shown that a projective plane of order 10 can have no symmetry, so that its construction, if it existed, would be complicated. An approach by means of the linear code associated with such a plane was suggested by MacWilliams et al. (1973). As a result of extensive computer searches the weight enumerator of the code (which would exist if a plane existed) was completely determined by Lam et al. (1986). Finally, at the end of 1988, Lam and his colleagues announced that a projective plane of order 10 does not exist (see Lam et al. 1989 and Lam 1991).

A recurrent problem in research on designs has been the construction of t-designs for large values of t. Although Kirkman (1853) had constructed 3-designs with parameters $(2^n, 4, 1)$, it soon became apparent that t-designs with $t > 3$ are hard to find (at least in the case when both repeated blocks and all k-subsets are forbidden). There are two 4-designs and two 5-designs associated with the *Mathieu groups*; the simplest, a 4-design with parameters (11, 5, 1) was given explicitly by Lea (1869). The others were known to Moore (1896) and were studied extensively by Carmichael (1931) and Witt (1938), but it was not until the 1960s that more 4-designs and 5-designs were discovered, by Alltop (1969), Assmus and Mattson (1969), and Denniston (1976). The suspicion that $t = 5$ might be an absolute limit was dispelled when Magliveras and Leavitt (1984) found several 6-designs with parameters (33, 8, 36). Soon afterwards, Teirlinck (1987) unfettered the combinatorial imagination by constructing t-designs (incomplete, and without repeated blocks) for all values of t. Recent results on designs are given by Jungnickel (1989).

7. Combinatorial set theory

If $m + 1$ objects are distributed among m pigeonholes, then at least one of the pigeonholes must receive at least two objects. This elementary fact, often called the *pigeonhole principle*, has been extensively generalized, giving rise in particular to that branch of combinatorics now known as *Ramsey theory*.

Although the pigeonhole principle is very well known, its origins are obscure. It appears in the literature as *Dirichlet's box principle*, and it was certainly used by Dirichlet in his study of the approximation of irrational numbers by rationals (Dirichlet 1879). However, the use of the pigeonhole principle certainly pre-dates Dirichlet. For example, Gauss used it in his *Disquisitiones Arithmeticae* (Gauss 1801), and it is likely that earlier uses of it occur in the literature.

The pigeonhole principle can be generalized in various directions. For example, if we distribute $m(k - 1) + 1$ objects among m pigeonholes, then at least one pigeonhole must receive at least k objects. A more profound and far-reaching generalization was given by Ramsey (1930). Whilst working on a problem in mathematical logic, he was led to the problem of distributing the r-subsets of an N-set into m classes in such a way that at least one class must contain every r-subset of some n-subset; this reduces to the generalized pigeonhole principle if $r = 1$, $n = k$, and $N = m(k - 1) + 1$. Ramsey's main achievement was to prove that, if N is sufficiently large, then any such distribution has this property. More generally, given integers $q_1, \ldots, q_m$, one can prove the existence of a number $R_r(q_1, \ldots, q_m)$ such that if $N \geqslant R_r(q_1, \ldots, q_m)$ and the r-subsets of an N-set are distributed in any manner into classes $C_1, \ldots, C_m$, then some class C_t must contain every r-subset of some q_t-subset; for example, $R_1(k, \ldots, k) = m(k - 1) + 1$. Ramsey also obtained an infinite version which asserts that if one distributes the r-subsets of an infinite set S into m classes, then some class must contain every r-subset of some infinite subset of S.

Five years later the subject was given a geometrical flavour by Erdős and Szekeres (1935). They observed that from any five points in general position in the plane one can always select four points forming a convex quadrilateral, and they generalized this by showing that, given enough points to start with, one can similarly form a convex n-gon. Their first proof of this result invoked Ramsey's theorem, but they also gave a combinatorial proof which depends on the fact that any sequence of $mn + 1$ distinct numbers must contain a decreasing sequence of length $m + 1$ or an increasing sequence of length $n + 1$; an eight-line proof of this latter result was later given by Seidenberg (1959).

Another well-known special case of Ramsey's theorem is to show that any gathering of six people must contain either three mutual acquaintances or three mutual non-acquaintances. More generally, we can ask for the minimum number $R_2(n, m)$ such that any gathering of $R_2(n, m)$ people must contain n mutual acquaintances or m mutual non-acquaintances. This can be expressed graphically by letting G be the graph of order $R_2(n, m)$ in which two vertices are joined by a red edge if the corresponding people are acquainted, and by a blue edge if not; then G must contain a red complete graph K_n or a blue complete graph K_m. In their 1935 paper, Erdős and Szekeres obtained the upper bound $(2n - 1)!/(m - 1)!^2$ for $R_2(n, m)$. Later, Erdős (1947) proved that $R_2(n, n) \geqslant 2^{n/2}$ and $R_2(3, n) \leqslant \frac{1}{2}n(n + 1)$. It is remarkable that these results have hardly been improved since then.

Two classical results which are also related to Ramsey's theorem are *Schur's lemma* and *Van der Waerden's theorem*. Schur (1916) proved that, if the integers $1, 2, \ldots, N$ are distributed into m classes, where $N > m!e$, then we can always find three integers x, y, z in some classes such that $x - y = z$. Analogues of this result for other linear equations, and for systems of linear equations, were later obtained by Rado (1943). An example of this is given by the system

$$x_1 - x_2 = x_2 - x_3 = \cdots = x_{k-1} - x_k .$$

If the x_i are unequal, then a solution of these equations consists of k integers in an arithmetic progression, and we deduce the result of Van der Waerden (1927) that if the integers $1, 2, \ldots, N$ are distributed into m classes, where N is sufficiently large compared with m, then at least one class must contain an arithmetic progression of any given size.

In recent years there has been an explosion of interest in Ramsey-type results, mainly through the influence of Erdős and his followers. An excellent source of recent results in Ramsey theory is Graham et al. (1980); see also chapter 25.

Another of the most influential results in combinatorial set theory is *Hall's theorem* (P. Hall 1935), sometimes called the *marriage theorem* (Halmos and Vaughan 1950), which gives a necessary and sufficient condition for a family $\mathrm{F} = (A_1, \ldots, A_m)$ of subsets of a set S to have a *transversal* (or *system of distinct representatives*); that is, a set of m distinct elements of S, one chosen from each of the sets A_i. Hall's theorem states that F has a transversal if and only if

$$\left| \bigcup_{i \in T} A_i \right| \geqslant T ,$$

for any subset T of $\{1, 2, \ldots, m\}$. If S has a matroid structure defined on it, then there is a corresponding condition, due to Rado (1942), for the existence of a transversal which is independent in the matroid. Later, M. Hall (1945) used the idea of a transversal to extend latin rectangles to latin squares.

Although Hall's Theorem is fundamental, other results related to it had been proved a few years earlier. The first of these was due to Frobenius (1912), and concerned the reducibility of the determinant of a matrix. A shorter proof of Frobenius's result, but in the language of bipartite graphs, was given by König (1915). In the following year, König (1916) gave a necessary and sufficient condition for a bipartite graph to contain a *perfect matching* (*1-factor*); this result was later extended by Tutte (1947), who gave a corresponding condition for an arbitrary graph to have a 1-factor. In 1917, Frobenius gave another proof of his theorem, using a lemma which is equivalent to Hall's theorem (Frobenius 1917). An excellent survey of these early results appears in Lovász and Plummer (1986). Matchings are discussed in chapter 3.

The above results have given rise to a large number of *minimax theorems* in combinatorics, in which the minimum of one quantity equals the maximum of another. Celebrated amongst these are *Menger's theorem* (Menger 1927), which states that the minimum number of vertices separating two given vertices in a graph is equal to the maximum number of vertex-disjoint paths between them, and *König's minimax theorem* (König 1931), that the size of a largest matching in a bipartite graph is equal to the smallest set of vertices which together touch every edge. Later, Ford and Fulkerson (1956), and independently Elias et al. (1956), proved the celebrated *max-flow min-cut theorem* for capacitated networks, which states that the maximum flow between two vertices is equal to the minimum capacity of a cut separating them. In another direction is *Dilworth's theorem* for partially ordered sets (Dilworth 1950), that the minimum number of *chains* (totally ordered sets) which cover a partially ordered set is equal to the maximum size of an *antichain* (set of incomparable elements). All such minimax results are related to the *duality theorem for linear programming*, and surveys of them can be found in Woodall (1978) and Schrijver (1983).

Another classical result of set theory is *Sperner's lemma* (Sperner 1928). This states that, if S is an n-set and F is a family of subsets of S none of which contains another (a *Sperner family*), then F contains at most

$$\binom{n}{[n/2]}$$

sets. This result has been extended by Lubell (1966) and others, who asserted that if $(A_1, \ldots, A_m)$ is a Sperner family, then

$$\sum_i \binom{n}{|A_i|}^{-1} \leq 1 .$$

A related result is the above theorem of Dilworth, since if P is the lattice of subsets of S, then an antichain in P is a Sperner family of S, and we can thus partition P into

$$\binom{n}{[n/2]}$$

chains. Also related is a result of Kleitman (1970), generalizing a problem of

Littlewood and Offord (1943), that if $x_1, \ldots, x_m$ are vectors in $\mathbb{R}^m$ with length at least 1, then there are at most

$$\binom{n}{[n/2]}$$

sums $x_j + \cdots + x_k$ which differ in norm by less than 1.

Instead of considering Sperner families, we can study intersecting families, in which any two sets have non-empty intersection. It is easy to see that if S is an n-set and $r > \frac{1}{2}n$, then any intersecting family of r-subsets of S has at most 2^{n-1} subsets. The problem is more difficult if $r \leq \frac{1}{2}n$, but is answered by the *Erdős–Ko–Rado theorem* (Erdős et al. 1961) which asserts that the maximum number of subsets is

$$\binom{n-1}{r-1},$$

and that this number is attained only when all the subsets contain a common element of S. The Erdős–Ko–Rado theorem has proved to be a milestone in extremal set theory.

The following problem was solved by Kruskal (1963), and independently by Katona: if F is any family of r-subsets of a finite S, what is the least number of $(r-1)$-subsets contained in some set in F? Further details of the result (now known as the *Kruskal–Katona theorem*), and of many other results in combinatorial set theory, can be found in Bollobás (1986).

8. Algorithmic combinatorics

Graph algorithms go back at least as far as the 1880s, when Fleury gave a method for tracing an Eulerian trail in a graph, and Trémaux and Tarry both showed how to traverse a maze (BLW 1D). It is in the 20th century that graph algorithms have come into their own, with the solutions of such problems as the shortest path problem, the minimum spanning-tree problem, and the Chinese postman problem. The *greedy algorithm* for finding a minimum-length spanning tree is often attributed to Kruskal (1956), but had been obtained some years earlier by Borůvka (1926). There are several algorithms for finding the shortest path in a network, of which the best known is due to Dijkstra (1959). Finding a longest path, or critical path, in an activity network dates from around the same time, with PERT (Program Evaluation and Review Technique) designed in the mid-1950s for a problem involving submarines. The *Chinese postman problem*, for finding the shortest route covering each edge of a graph, was solved by Guăn (= 管 1960). Some of these, and other graph algorithms, are discussed in chapters 28 and 35.

A related problem is the *travelling salesman problem*, in which a salesman has to make a cyclic tour of a number of cities in minimum time or distance. A rudimentary statement of the problem appeared in a practical German book written for the *Handlungsreisende* (Voigt 1831), but its first appearance in mathematical circles took place in the early 1930s at Princeton. Its main publicist was Flood, who later popularized the problem at the RAND Corporation. This led eventually to the fundamental paper of Dantzig et al. (1954) involving the solution of a travelling salesman problem with 49 cities. Over the years the number of cities considered has been gradually increased and in 1986 a problem with 2392 cities was settled by Padberg and Rinaldi (1987). An extensive treatment of the travelling salesman problem appears in Lawler et al. (1985).

A substantial advance in the understanding of combinatorial algorithms was the classification of combinatorial problems as "easy" or "hard". By the late 1960s it was already clear that problems such as the travelling salesman problem seemed to be much more difficult than, for example, the minimum spanning-tree problem. Edmonds (1965) had already described an algorithm as "good" if a polynomial-time algorithm exists, and in three fundamental papers, Cook (1971), Karp (1972) and Levin (1973) developed the concept of *NP-completeness*. In this theory the assignment, transportation and minimum spanning-tree problems are all in the *polynomial class P*, whereas the travelling salesman and Hamiltonian cycle problems are *NP-hard*. The concept of computational complexity is discussed in chapter 29.

The travelling salesman problem was not the only significant combinatorial problem studied at the RAND Corporation in the 1940s and 1950s. During this time techniques were developed, by Dantzig and Fulkerson (1954) for finding the least number of tankers needed to meet a fixed schedule, by Ford and Fulkerson (1956) for finding the maximum flow in a capacitated network, and by Gomory and Hu (1961) for investigating multi-terminal and multi-commodity flows. These investigations led eventually to the subject of *polyhedral combinatorics*, as described in detail in chapters 28 and 30.

Before joining the RAND Corporation in 1952, Dantzig had instigated the study of *linear programming* techniques. The basic ideas of linear programming can be traced back to Fourier (1826); an account of Fourier's study of systems of linear inequalities, with some historical remarks, is given by Williams (1986). Later, in the 1930s, Kantorovitch considered linear programming as a mathematical study in its own right, but his work remained unnoticed for many years. In the 1940s, motivated by Second World War planning activities, Dantzig and von Neumann independently discovered and developed the idea of linear programming. Dantzig proposed the basic theory in 1947–8 (Dantzig 1949), and the fundamental concept of duality was introduced by von Neumann in 1947. Later, Dantzig introduced the highly efficient and practical *simplex method* for solving linear programming problems (Dantzig 1951). Further information about the origins of linear programming problems and their connections with matching theory can be found in Lovász and Plummer (1986) and in a historical article by Dantzig (1982). Another good source which includes historical material is the book by Schrijver (1986).

Linear programming techniques proved to be ideally suited to the solution of certain practical problems which had arisen during the war years. In particular, the fundamental *diet problem* of selecting a diet with maximum nutritional value was discussed by Stiegler (1945), and the *transportation problem* of shipping a commodity at minimum cost from several sources to several markets was investigated by Hitchcock (1941), although it had been studied geometrically 160 years earlier by Monge (1784).

Related to the above topics is the earlier creation of the theory of two-person games by von Neumann (1928). This paper contained the fundamental *minimax theorem* for games, although the proof there is involved. Simpler proofs, and an extensive treatment of the subject in general, were later given in the pioneering books of von Neumann and Morgenstern (1944) and McKinsey (1952).

Since the 1940s, the topics mentioned very briefly in this section have grown in importance, and now play a central role in modern combinatorics. Further information about most of these subjects can be found elsewhere in this Handbook, in particular in chapters 28, 30 and 35.

References

Ahmedev, S.S., and B.A. Rosenfeld

[1963] Al-Tusi's arithmetic (in Russian), *Istor.-Mat. Issled.* **15**, 431–444.

Alltop, W.O.

[1969] An infinite class of 4-designs, *J. Combin. Theory* **6**, 320–322.

Andrews, G.E.

[1976] *The Theory of Partitions, Encyclopedia of Mathematics*, Vol. 2 (Addison-Wesley, Reading MA).

[1979] Book review, *Bull. Amer. Math. Soc. (New Ser.)* **1**, 989 997.

Appel, K., and W. Haken

[1977] The solution of the four-color-map problem, *Sci. Amer.* **237**(4), 108–121.

[1989] *Every Planar Map is Four Colorable* (American Mathematical Society, Providence, RI).

Arbogast, L.-A.-F.

[1800] *Du Calcul des Dérivations* (Levrault, Strasbourg).

Assmus, E.F., and H.F. Mattson

[1969] New 5-designs, *J. Combin. Theory* **6**, 122–151.

Bell, E.T.

[1938] The history of Blissard's symbolic method, with a sketch of its inventor's life, *Amer. Math. Monthly* **45**, 414–421.

Bermond, J.-C.

[1978] Hamiltonian graphs, in: *Selected Topics in Graph Theory*, eds. L.W. Beineke and R.J. Wilson (Academic Press, London) pp. 127–167.

Biggs, N.L.

[1979] The roots of combinatorics, *Historia Math.* **6**, 109 136.

[1981] T.P. Kirkman, mathematician, *Bull. London Math. Soc.* **13**, 97–120.

[1983] De Morgan on map colouring and the separation axiom, *Arch. Hist. Exact Sci.* **28**, 165–170.

Biggs, N.L., E.K. Lloyd and R.J. Wilson

[1976] *Graph Theory 1736–1936* (Clarendon Press, Oxford). Revised edition: 1986.

Binet, J.P.M.

[1812] Mémoire sur un système de formules analytiques et leur application à des considérations géometriques, *J. École Polytech.* **9**, 280–302.

Birkhoff, G.D.

[1912–13] A determinant formula for the number of ways of coloring a map, *Ann. of Math. (2)* **14**, 42–46 [= *Math. Papers*, Vol. 3, pp. 1–5].

[1913] The reducibility of maps, *Amer. J. Math.* **35**, 115–128 [= *Math. Papers*, Vol. 3, pp. 6–19].

Bollobás, B.

[1986] *Combinatorics* (Cambridge University Press, London).

Borůvka, O.

[1926] O jistém problému minimálnim, *Acta Soc. Sci. Natur. Moravicae* **3**, 37–58.

Bose, R.C.

[1938] On the application of the properties of Galois fields to the construction of hyper-Graeco-Latin squares, *Sankhyā* **3**, 323–338.

[1939] On the construction of balanced incomplete block designs, *Ann. Eugenics* **9**, 353–399.

Bose, R.C., and S.S. Shrikhande

[1959] On the falsity of Euler's conjecture about the non-existence of two orthogonal latin squares of order $4t+2$, *Proc. Nat. Acad. Sci. U.S.A.* **45**, 734–737.

Bose, R.C., S.S. Shrikhande and E.T. Parker

[1960] Further results on the construction of mutually orthogonal latin squares and the falsity of Euler's conjecture, *Canad. J. Math.* **12**, 189–203.

Brooks, R.L.

[1941] On colouring the nodes of a network, *Proc. Cambridge Philos. Soc.* **37**, 194–197.

Bruck, R.H., and H.J. Ryser

[1949] The non-existence of certain finite projective planes, *Canad. J. Math.* **1**, 88–93.

Carmichael, R.D.

[1931] Tactical configurations of rank two, *Amer. J. Math.* **53**, 217–240.

Cartier, P., and D. Foata

[1969] *Problèmes Combinatoires de Commutation et Réarrangements, Lecture Notes in Mathematics*, Vol. 85 (Springer, Berlin).

Cauchy, A.

[1812] Mémoire sur les fonctions qui ne peuvent obtenir que deux valeurs égales et de signes contraires par suite des transpositions opérées entre les variable qu'elles renferment, *J. École Polytech.* **10**, 29–112.

[1815] Sur les fonctions qui ne peuvent obtenir que deux valeurs égales, *J. École Polytech.* **17**, 90–169.

Chowla, A.S., and H.J. Ryser

[1950] Combinatorial problems, *Canad. J. Math.* **2**, 93–99.

Cook, S.A.

[1971] The complexity of theorem-proving procedures, in: *Proc. 3rd Annu. ACM Symp. Theory of Computing, Shaker Heights, OH, 1971* (ACM, New York) pp. 151–158.

Coupy, É.

[1851] Solution d'un problème appartenant à la géométrie de situation par Euler, *Nouv. Ann. Math.* **10**, 106–119.

Dantzig, G.B.

[1949] Programming in a linear structure, *Econometrica* **17**, 73–74.

[1951] Maximization of linear functions of variables subject to linear inequalities, in: *Activity Analysis of Production and Allocation*, ed. T.C. Koopmans (Wiley, New York) pp. 339–347.

[1982] Reminiscences about the origins of linear programming, *Oper. Res. Lett.* **1**, 43–48. Also published 1983, in: *Mathematical Programming – The State of the Art, Bonn*, eds. A. Bachem, M. Grötschel and B. Korte (Springer, Berlin) pp. 78–86.

Dantzig, G.B., and D.R. Fulkerson

[1954] Minimizing the number of tankers to meet a fixed schedule, *Naval Res. Logist. Quart.* **1**, 217–222.

Dantzig, G.B., D.R. Fulkerson and S.M. Johnson

[1954] Solution of a large-scale traveling-salesman problem, *Oper. Res.* **2**, 393–410.

David, F.N.

[1962] *Games, Gods and Gambling* (Griffin, London).

David, F.N., M.G. Kendall and D.E. Barton

[1966] *Symmetric Function and Allied Tables* (Cambridge University Press, Cambridge).

De Moivre, A.

[1697] A method of raising an infinite multinomial to any given power, or extracting any given root of the same, *Philos. Trans.* No. 230, 619–625.

[1698] A method of extracting the root of an infinite equation, *Philos. Trans.* No. 240, 190–193.

[1718] *The Doctrine of Chances* (private printing, London).

[1722] De fractionibus algebraicis radicalitate immunibus ad fractiones simpliciores reducendis, deque summandis terminus quarumdam serierum aequali intervallo a se distantibus, *Philos. Trans.* No. 373, 162–178.

[1730] *Miscellanea Analytica de Seriebus et Quadraturis* (private printing, London) pp. 26–42.

de Vries, H.L.

[1984] Historical notes on Steiner systems, *Discrete Math.* **52**, 293–297.

Denniston, R.H.F.

[1976] Some new 5-designs, *Bull. London Math. Soc.* **8**, 263–267.

Dijkstra, E.W.

[1959] A note on two problems in connexion with graphs, *Numer. Math.* **1**, 269–271.

Dilworth, R.P.

[1950] A decomposition theorem for partially ordered sets, *Ann. of Math. (2)* **51**, 161–166.

Dirac, G.A.

[1952] Some theorems on abstract graphs, *Proc. London Math. Soc. (3)* **2**, 69–81.

Dirichlet, P.G.L.

[1879] *Vorlesungen über Zahlentheorie* (Vieweg, Braunschweig).

Durfee, W.P.

[1882/1883] On the number of self-opposite partitions, *Johns-Hopkins Univ. Circulars* **2**, 23.

Dutka, J.

[1986] On the problème des ménages, *Math. Intelligencer* 8, 18–25 & 33.

Edmonds, J.R.

[1965] Paths, trees and flowers, *Canad. J. Math.* **17**, 449–467.

Edmonds, J.R., and D.R. Fulkerson

[1965] Transversals and matroid partitions, *J. Res. Nat. Bur. Standards B* **69**, 147–153.

Edwards, A.W.F.

[1987] *Pascal's Arithmetical Triangle* (Charles Griffin/Oxford University Press, London/New York).

Elias, P., A. Feinstein and C.E. Shannon

[1956] Note on maximum flow through a network, *IRE Trans. Inform. Theory* **IT-2**, 117–119.

Erdős, P.

[1938] On sequences of integers no one of which divides the product of two others and on some related problems, *Izv. Nauc. Isz Inszt. Mat. Mech. Tomsk* **2**, 74–82.

[1947] Some remarks on the theory of graphs, *Bull. Amer. Math. Soc.* **53**, 292–294.

Erdős, P., and A. Rényi

[1960] On the evolution of random graphs, *Magyar Tud. Akad. Mat. Kut. Int. Közl* **5A**, 17–61.

Erdős, P., and G. Szekeres

[1935] A combinatorial problem in geometry, *Comp. Math.* **2**, 463–470.

Erdős, P., C. Ko and R. Rado

[1961] Intersection theorems for systems of finite sets, *Quart. J. Math. Oxford (2)* **12**, 313–320.

Euler, L.

[1736] Solutio problematis ad geometriam situs pertinentis, *Comm. Acad. Sci. Imp. Petropol.* **8**, 128–140 [= *Opera Omnia (1)*, Vol. 7, 1–10].

[1748] *Introductio in Analysin Infinitorum*, Vol. 1 (Bousquet, Lausanne) ch. 16 [= *Opera Omnia (1)*, Vol. 8, 313–338]. English edition: 1988, *Introduction to Analysis of the Infinite* (Springer, New York).

[1751] Decouverte d'une loi tout extraordinaire des nombres par support à la somme de leurs diviseurs, *Bib. Impartiale* **3**, 10–31 [= *Opera Omnia (1)*, Vol. 2, 241–253].

[1759] Solution d'une question curieuse que ne paroit soumise à aucune analyse, *Mém. Acad. Sci. Berlin.* **15**, 310–337 [= *Opera Omnia (1)*, Vol. 7, 26–56].

[1782] Recherches sur une nouvelle espèce de quarrés magiques, *Verh. Zeeuwsch Genootsch. Wetensch. Vlissingen* **9**, 85–239 [= *Opera Omnia (1)*, Vol. 7, 291–392].

Fano, G.

[1892] Sui postulari fondamenti della geometria proiettiva, *Giorn. Mat.* **30**, 106–132.

Fisher, R.A.

[1940] An examination of the different possible solutions of a problem in incomplete blocks, *Ann. Eugenics* **10**, 52–75.

Fisher, R.A., and F. Yates

[1934] The 6×6 latin squares, *Proc. Cambridge Philos. Soc.* **30**, 492–507.

[1938] *Statistical Tables for Biological, Agricultural and Medical Research* (Oliver and Boyd, Edinburgh).

Fleischner, H.

[1974] The square of every two-connected graph is Hamiltonian, *J. Combin. Theory B* **16**, 29–34.

Ford, L.R., and D.R. Fulkerson

[1956] Maximal flow through a network, *Canad. J. Math.* **8**, 399–404.

Fourier, J.B.J.

[1826] Solution d'une question particulière du calcul des inégalités, *Œuvres*, Vol. 2, Paris, pp. 317–328.

Franklin, F.

[1881] Sur le développement du produit infini $(1-x)(1-x^2)(1-x^3)(1-x^4)\cdots$, *C.R. Acad. Sci. (Paris)* **92**, 448–450.

Franklin, P.

[1922] The four color problem, *Amer. J. Math.* **44**, 225–236.

Frobenius, G.

[1903] Über die charakteristischen Einheiten der symmätrischen Gruppe, *Sitzungsber. Königl. Preuß. Akad. Wiss. Berlin*, pp. 328–358 [= *Gesammelte Abhandlungen*, Vol. 3, pp. 244–274].

[1912] Über Matrizen aus nicht negativen Elementen, *Sitzungsber. Königl. Preuß. Akad. Wiss. Berlin* **26**, 456–477 [= *Gesammelte Abhandlungen*, Vol. 3, pp. 546–567].

[1917] Über zerlegbare Determinanten, *Sitzungsber. Königl. Preuß. Akad. Wiss. Berlin* **18**, 274–277 [= *Gesammelte Abhandlungen*, Vol. 3, pp. 701–704].

Gauss, C.F.

[1801] *Disquisitiones Arithmeticae* (Fleischer, Leipzig). English edition: 1966 (Yale University Press, New Haven, CT).

[1867] *Werke*, Vol. 5 (Göttingen).

Glaisher, J.W.L.

[1875] Formulae of verification in partitions, *Rep. Br. Assoc. Adv. of Sci.* **45**, 11–12.

[1909] On the number of partitions of a number into a given number of parts, *Quart. J. Pure Appl. Math.* **40**, 57–143.

Glover, H.H., J.P. Huneke and C.S. Wang

[1979] 103 graphs that are irreducible for the projective plane, *J. Combin. Theory B* **27**, 332–370.

Gomory, R.E., and T.C. Hu

[1961] Multi-terminal network flows, *SIAM J. Appl. Math.* **9**, 551–556.

Graham, R.L., B.L. Rothschild and J.H. Spencer

[1980] *Ramsey Theory* (Wiley, New York).

Guăn Méigŭ

[*see* Kwan Mei-Ku]

Hadwiger, H.

[1943] Über eine Klassifikation der Streckenkomplexe, *Vierteljschr. Naturforsch. Ges. Zürich* **88**, 133–142.

Hall, M.

[1945] An existence theorem for Latin squares, *Bull. Amer. Math. Soc.* **51**, 381–388.

[1947] Cyclic projective planes, *Duke Math. J.* **14**, 1079–1090.

Hall, P.

[1935] On representatives of subsets, *J. London Math. Soc.* **10**, 26–30.

Halmos, P.R., and H.E. Vaughan

[1950] The marriage problem, *Amer. J. Math.* **72**, 214–215.

Hammond, J.

[1882] On the calculation of symmetric functions, *Proc. London Math. Soc.* **13**, 79–84.

[1883] On the use of certain differential operators in the theory of equations, *Proc. London Math. Soc.* **14**, 119–129.

Hardy, G.H., and S. Ramanujan

[1918] Asymptotic formulae in combinatory analysis, *Proc. London Math. Soc. (2)* **17**, 75–115.

Herschel, J.F.W.

[1850] On the algebraic expression of the number of partitions of which a given number is susceptible, *Philos. Trans. Roy. Soc. London* **140**, 399–422.

Hindenburg, C.F.

[1796] *Sammlung Combinatorisch–Analytischer Abhandlungen*, 2 volumes (Fleischer, Leipzig).

Hitchcock, F.L.

[1941] The distribution of a product from several sources to numerous localities, *J. Math. Phys.* **20**, 224–230.

Hughes, B.

[1989] The arithmetical triangle of Jordanus de Nemore, *Historia Math.* **16**, 213–223.

Jungnickel, D.

[1989] Design theory: an update, *Ars Combin.* **29**, 129–199.

Karp, R.M.

[1972] Reducibility among combinatorial problems, in: *Complexity of Computer Computations*, eds. R.E. Miller and J.W. Thatcher (Plenum Press, New York) pp. 85–103.

Kennedy, J.W., L.V. Quintas and M.M. Sysło

[1985] The theorem on planar graphs, *Historia Math.* **12**, 356–368.

Kirkman, T.P.

[1847] On a problem in combinations, *Cambridge and Dublin Math. J.* **2**, 191–204.

[1853] Theorems on combinations, *Cambridge and Dublin Math. J.* **8**, 38–45.

Kleitman, D.J.

[1970] On a lemma of Littlewood and Offord on the distribution of linear combinations of vectors, *Adv. in Math.* **5**, 1–3.

Knobloch, E.

[1974] Leibniz on combinatorics, *Historia Math.* **1**, 409–430.

[1979] Musurgia universalis: unknown combinatorial studies in the age of baroque absolutism, *Hist. Sci.* **17**, 258–275.

König, D.

[1915] Vonalrendszerek és determinánsok (Line systems and determinants), *Math. Termész. Ért.* **33**, 221–229.

[1916] Gráfok és alkalmazásuk a determinánsok és halmazok elméletében, *Math. Termész. Ért.* 34, 104–119 [Über Graphen und ihre Anwendung auf Determinantentheorie und Mengenlehre, *Math. Ann.* 77, 453–465].

[1927] Über eine Schlussweise aus dem Endlichen ins Unendliche (Punktmengen – Kartenfärben – Verwandtschaftsbeziehungen – Schachspiel), *Acta Litt. Sci. Szeged* **3**, 121–130.

[1931] Graphok és matrixok, *Mat. Fiz. Lapok* **38**, 116–119.

[1936] *Theorie der Endlichen und Unendlichen Graphen* (Akademische Verlagsgesellschaft, Leipzig). Reprinted: 1950 (Chelsea, New York). English edition: 1990, *Theory of Finite and Infinite Graphs* (Birkhäuser, Boston).

Kruskal, J.B.

[1956] On the shortest spanning subtree of a graph and the traveling salesman problem, *Proc. Amer. Math. Soc.* **7**, 48–50.

[1963] The number of simplices in a complex, in: *Mathematical Optimization Techniques*, ed. R. Bellman (University of California Press, Berkeley, CA) pp. 251–278.

Kung, J.P.S., and G.-C. Rota

[1984] The invariant theory of binary forms, *Bull. Amer. Math. Soc. (New Ser.)* **10**, 27–85.

Kwan Mei-Ku (Guăn Méigŭ)

[1960] Graphic programming using odd or even points, *Acta Math. Sinica* **10**, 263–266 [1962, *Chinese Math.* **1**, 273–277].

Lagrange, J.-L. (de la Grange)

[1770] Nouvelle méthode pour résoudre les équations littérales par le moyen des séries, *Hist. Acad. Royale Sci. Belles-Lettres Berlin* **24** [= 1869, *Œuvres*, Vol. 3, ed. J.A. Serret (Gauthier-Villars, Paris) pp. 5–73]. Reprinted: 1973 (G. Ohms, Hildesheim).

Lam, C.W.H.

[1991] The search for a finite projective plane of order 10, *Amer. Math. Monthly* **98**, 305–318.

Lam, C.W.H., L. Thiel and S. Swiercz

[1986] The non-existence of code words of weight 16 in a projective plane of order 10, *J. Combin. Theory A* **42**, 207–214.

[1989] The non-existence of finite projective planes of order 10, *Can. J. Math.* **61**, 1117–1123.

Lawler, E.L., J.K. Lenstra, A.H.G. Rinnooy Kan and D.B. Shmoys

[1985] eds., *The Traveling Salesman Problem: A Guided Tour through Combinatorial Optimization* (Wiley, Chichester).

Lea, W.

[1869] Solution to question 2244, *Math. Quest. Educ. Times* **9**, 35–36.

Legendre, A.M.

[1794] *Éléments de Géométrie*, 1st Ed. (Firmin Didot, Paris).

Leibniz, G.W.

[1666] *Dissertatio de Arte Combinatoria* (Fickium & Seuboldum, Leipzig) [= *Mathematische Schriften*, Vol. 5, pp. 7–79].

Leonardo of Pisa [= Fibonacci]

[1202] *Liber Abaci* [= 1957, *Scritti di Leonardo Pisano*, Vol. 1 (Rome) pp. 283–284].

Levin, L.A.

[1973] Universal sequential search problems, *Problemy Peridachi Informatsii* 9, 115–116 [1984, *Ann. Hist. Comput.* **6**, 399-400].

Li Yan, and Du Shiran

[1987] *Chinese Mathematics: a Concise History* (Clarendon Press, Oxford).

Listing, J.B.

[1847] Vorstudien zur Topologie, *Göttinger Studien (Abt. I) Math. Naturwiss. Abh.* **1**, 811–875. Printed separately as a book: 1848 (Göttingen).

[1861–62] Der Census räumlicher Complexe oder Verallgemeinerung des Euler'schen Satzes von der Polyëdern, *Abh. K. Ges. Wiss. Göttingen Math. Cl.* **10**, 97–182. Printed separately as a book: 1862 (Dieterich'sche Verlagshandlung, Göttingen).

Littlewood, J.E., and A.C. Offord

[1943] On the number of real roots of a random algebraic equation III, *Mat. Sb.* **12**, 277–285.

Lloyd, E.K.

[1984] J. Howard Redfield 1879–1944, *J. Graph. Theory* **8**, 195–203.

[1988] Redfield's papers and their relevance to counting isomers and isomerizations, *Discrete Math.* **19**, 289–304. Reprinted in: *Applications of Graphs in Chemistry and Physics*, eds. J.W. Kennedy and L.V. Quintas (North-Holland, Amsterdam) pp. 289–304.

Lovász, L., and M.D. Plummer

[1986] *Matching Theory, Ann. Discrete Math.* **29**.

Lubell, D.

[1966] A short proof of Sperner's lemma, *J. Combin. Theory* **1**, 299.

Lucas, É.

[1877] Théorie nouvelle des nombres de Bernoulli et d'Euler, *Ann. Mat. Pura et Appl. (2)* **8**, 56–79.

[1882] *Récréations Mathématiques*, Vol. 1 (Gauthier-Villars, Paris).

Lunn, A.C., and J.K. Senior

[1929] Isomerism and configuration, *J. Phys. Chem.* **33**, 1027–1079.

MacInnes, C.R.
[1907] Finite planes with less than eight points on a line, *Amer. Math. Monthly* **14**, 171–174.
MacMahon, P.A.
[1896–97] Combinatory analysis: a review of the present state of knowledge, *Proc. London Math. Soc.* **28**, 5–32 [= *Collected Papers*, Vol. 2, ch. 19].
[1915–16] *Combinatory Analysis*, 2 volumes (Cambridge University Press, Cambridge).
[1978/86] *Collected Papers*, 2 volumes, ed. G.E. Andrews (MIT Press, Cambridge, MA).
MacNeish, H.F.
[1922] Euler squares, *Ann. of Math. (2)* **23**, 221–227.
MacWilliams, F.J., N.J.A. Sloane and J.W. Thompson
[1973] On the existence of a projective plane of order 10, *J. Combin. Theory A* **14**, 66–78.
Magliveras, S.S., and D.W. Leavitt
[1984] Simple 6-(33,8,36) designs from $P\Gamma L_2(32)$, in: *Computational Group Theory*, ed. M.D. Atkinson (Academic Press, London) pp. 337–351.
McKinsey, J.C.C.
[1952] *Introduction to the Theory of Games, The RAND series* (McGraw-Hill, New York).
Menger, K.
[1927] Zur allgemeinen Kurventheorie, *Fund. Math.* **10**, 96–115.
Minc, H.
[1978] *Permanents, Encyclopedia of Mathematics*, Vol. 6 (Addison-Wesley, Reading, MA).
[1983] Theory of permanents 1978–1981, *Linear Algebra and Multi. Algebra* **12**, 227–263.
Monge, G.
[1784] Mémoire sur la théorie des déblais et des remblais, *Hist. Acad. Royale Sci. avec Mém. Math. et Phys.* (année 1781) 2e partie, *Histoire* 34–38, *Mémoire* 666–704.
Moore, E.H.
[1896] Tactical memoranda I–III, *Amer. J. Math.* **18**, 264–303.
Neumann, P.M.
[1979] A lemma that is not Burnside's, *Math. Sci.* **4**, 133–141.
Nicholson, P.
[1818] *Essays on the Combinatorial Analysis* (Longman et al., London).
Ore, O.
[1960] Note on Hamiltonian circuits, *Amer. Math. Monthly* **67**, 55.
Oxley, J.G.
[1992] *Matroid Theory* (Oxford University Press, New York).
Ozanam, J.
[1725] *Récréations Mathématiques et Physiques*, New Edition (Jombert, Paris).
Padberg, M.W., and G. Rinaldi
[1987] Optimization of a 532-city symmetric traveling salesman problem by branch and cut, *Oper. Res. Lett.* **6**, 1–7.
Paley, R.E.A.C.
[1933] On orthogonal matrices, *J. Math. Phys.* **12**, 311–320.
Parker, E.T.
[1959a] Construction of some sets of mutually orthogonal latin squares, *Proc. Amer. Math. Soc.* **10**, 946–949.
[1959b] Orthogonal latin squares, *Proc. Nat. Acad. Sci. U.S.A.* **45**, 859–862.
Pascal, B.
[1665] *Traité du Triangle Arithmétique avec Quelques Autres Petits Traités sur la Mesure Matière* (Desprez, Paris).
Poinsot, L.
[1810] Sur les polygones et les polyèdres, *J. École Polytech.* 4(Cah. 10), 16–48.
Pólya, G.
[1937] Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen, *Acta Math.* **68**, 308–416.

Pólya, G., and R.C. Read

[1987] *Combinatorial Enumeration of Groups, Graphs and Chemical Compounds* (Springer, New York).

Pont, J.C.

[1974] *La Topologie Algébrique des Origines à Poincaré* (Presses Universitaires de France, Paris).

Rado, R.

[1942] A theorem on independence relations, *Quart J. Math. (Oxford Ser.)* **13**, 83–89.

[1943] Note on combinatorial analysis, *Proc. London Math. Soc. (2)* **48**, 122–160.

Ramanujan, S.

[1919] Some properties of *p(n)*, the number of partitions of *n*, *Proc. Cambridge Philos. Soc.* **19**, 207–210 [= *Collected Papers*, pp. 210–213].

Ramsey, F.P.

[1930] On a problem of formal logic, *Proc. London Math. Soc. (2)* **30**, 264–286.

Ray-Chaudhuri, D.K., and R.M. Wilson

[1971] Solution of Kirkman's schoolgirl problem, *Proc. Symp. Pure Math.* **19**, 187–203.

Read, R.C.

[1959] The enumeration of locally restricted graphs, *J. London Math. Soc.* **34**, 417–436.

[1968] The use of S-functions in combinatorial analysis, *Canad. J. Math.* **20**, 808–841.

Redfield, J.H.

[1927] The theory of group-reduced distributions, *Amer. J. Math.* **49**, 433–455.

Ringel, G.

[1974] *Map Color Theorem* (Springer, Berlin).

Robertson, N., and P.D. Seymour

[1985] Graph minors – a survey, in: *Surveys in Combinatorics 1985, London Mathematical Society Lecture Note Series*, Vol. 103, ed. I. Anderson (Cambridge University Press, Cambridge) pp. 153–171.

Robinson, G. de B.

[1938] On the representations of the symmetric group I, *Amer. J. Math.* **60**, 745–760.

Roman, S.

[1984] *Umbral Calculus* (Academic Press, Orlando, FL).

Rota, G.-C.

[1964] On the foundations of combinatorial theory I. Theory of Möbius functions, *Z. Wahrscheinlichkeitstheorie u. Verw. Gebiete* **2**, 340–368.

Rouse Ball, W.W.

[1892] *Mathematical Recreations and Problems of Past and Present Times* (Macmillan, London).

Sachs, H., M. Stiebitz and R.J. Wilson

[1988] An historical note: Euler's Königsberg letters, *J. Graph Theory* **12**, 133–139.

Schensted, C.

[1961] Longest increasing and decreasing subsequences, *Canad. J. Math.* **13**, 179–191.

Schneider, I.

[1968–69] Der Mathematiker Abraham de Moivre, *Arch. Hist. Exact Sci.* **5**, 177–317.

Schrijver, A.

[1983] Min–max results in combinatorial optimization, in: *Mathematical Programming – The State of the Art, Bonn*, eds. A. Bachem, M. Grötschel and B. Korte (Springer, Berlin) pp. 439–500.

[1986] *Theory of Linear and Integer Programming* (Wiley, Chichester).

Schur, I.

[1916] Über die Kongruenz $x^m+y^m = z^m \pmod{p}$, *Jber Deutsch. Math.-Verein.* **25**, 114–117.

Seidenberg, A.

[1959] A simple proof of a theorem of Erdős and Szekeres, *J. London Math. Soc.* **34**, 352.

Singer, J.

[1938] A theorem in finite projective geometry and some applications to number theory, *Trans. Amer. Math. Soc.* **43**, 377–385.

Sperner, E.

[1928] Ein Satz über Untermengen einer endlichen Menge, *Math. Z.* **27**, 544–548.

Steiner, J.

[1853] Combinatorisches Aufgabe, *J. Reine und Angew. Math.* **45**, 181–182.

Stevens, W.L.

[1939] The completely orthogonalized Latin square, *Ann. Eugenics* **9**, 82 93.

Stiegler, G.J.

[1945] The cost of subsistence, *J. Farm. Econ.* **27**, 303 314.

Street, A.P., and D.J. Street

[1988] Latin squares and agriculture: the other bicentennial, *Math. Sci.* **13**, 48-55.

Sylvester, J.J.

[1855–57] On the partitions of numbers, *Quart. J. Pure Appl. Math.* **1**, 141–152 [= *Math. Papers*, Vol. 2, pp. 90 99].

[1878] On an application of the new atomic theory to the graphical representation of the invariants and covariants of binary quantics – with three appendices, *Amer. J. Math.* **1**, 64–125 [= *Math. Papers*, Vol. 3, pp. 148-206].

Takács, L.

[1981] On the problème des ménages, *Discrete Math.* **36**, 289–298.

Tarry, G.

[1900] Le problème des 36 officiers, *C.R. Assoc. France Av. Sci.* **29**, 170–203.

Teirlinck, L.

[1987] Non-trivial t-designs without repeated blocks exist for all t, *Discrete Math.* **65**, 301–311.

Thomassen, C.

[1983] Infinite graphs, in: *Selected Topics in Graph Theory 2*, eds. L.W. Beineke and R.J. Wilson (Academic Press, London) pp. 129–160.

Todd, J.A.

[1933] A combinatorial problem, *J. Math. Phys.* **12**, 321–333.

Turán, P.

[1941] On an extremal problem in graph theory (in Hungarian), *Mat. Fiz. Lapok* **48**, 436 452.

Tutte, W.T.

[1947] The factorisation of linear graphs, *J. London Math. Soc.* **22**, 107–111.

[1959] Matroids and graphs, *Trans. Amer. Math. Soc.* **90**, 527–552.

van der Waerden, B.L.

[1927] Beweis einer Baudetschen Vermutung, *Nieuw Arch. Wisk.* **15**, 212–216.

Veblen, O., and W.H. Bussey

[1906] Finite projective geometries, *Trans. Amer. Math. Soc.* **7**, 241–259.

Veblen, O., and J.H.M. Wedderburn

[1907] Non-Desarguesian and non-Pascalian geometries, *Trans. Amer. Math. Soc.* **8**, 379–388.

Vizing, V.G.

[1964] On an estimate of the chromatic class of a p-graph (in Russian), *Diskret. Analiz* **3**, 25 30.

[1965] The chromatic class of a multigraph, *Cybernetics* **1**(3), 32–41.

Voigt, B.F.

[1831] *Der Handlungsreisende, wie er sein soll und was er zu thun hat, um Aufträge zu erhalten und einer glücklichen Erfolgs in seinen Geschäften gewiss zu sein (von einem alten Commis-Voyageur, Ilmenau)*. Reprinted: 1981 (Schramm, Kiel).

von Neumann, J.

[1928] Zur Theorie der Gesellschaftsspiele, *Math. Ann.* **100**, 295–320.

von Neumann, J., and O. Morgenstern

[1944] *Theory of Games and Economic Behavior* (Princeton University Press, Princeton, NJ).

von Staudt, K.G.C.

[1856] *Beiträge zur Geometrie der Lage*, 3 volumes (Korn'schen Buchhandlung, Nürnberg).

Warburton, H.

[1842–49] On the partitions of numbers, and on combinations and permutations, *Trans. Cambridge Philos. Soc.* **8**, 471–492.

Welsh, D.J.A.

[1976] *Matroid Theory* (Academic Press, London).

White, A.T.

[1983] Ringing the changes, *Math. Proc. Cambridge Philos. Soc.* **94**, 203–215.

Whitney, H.

[1931] Non-separable and planar graphs, *Proc. Nat. Acad. Sci. U.S.A.* **17**, 125–127.

[1935] On the abstract properties of linear dependence, *Amer. J. Math.* **57**, 509–533.

Williams, H.P.

[1986] Fourier's method of linear programming and its dual, *Amer. Math. Monthly* **93**, 681–695.

Wilson, R.M.

[1975] An existence theory for pairwise balanced designs, III: proof of the existence conjecture, *J. Combin. Theory A* **18**, 71–79.

Witt, E.

[1938] Über Steinersche Systeme, *Abh. Math. Sem. Hamburg* **12**, 265–275.

Woodall, D.R.

[1978] Minimax theorems in graph theory, in: *Selected Topics in Graph Theory*, eds. L.W. Beineke and R.J. Wilson (Academic Press, London) pp. 237–269.

Wu, L., L. Zhu and Q. Kang

[1990] eds., *Collected Works of Lu Jiaxi on Combinatorial Designs* (Inner Mongolia People's Press, Huhhot).

Wythoff, W.A.

[1906–10] A solution to problem 28, *Wisk. Opg. Wisk. Genoot. (New Ser.)* **10**, 60–61.

Yates, F.

[1936] Incomplete randomized blocks, *Ann. Eugenics* **7**, 121–140. Reprinted: 1970, in: *Experimental Design: Selected Papers of Frank Yates*, eds. W.G. Cochran, D.J. Finney and M.J.R. Healey (Griffin, London).

Young, A.

[1901] Quantitative substitutional analysis I, *Proc. London Math. Soc.* **33**, 97–146.

[1977] *The Collected Papers of Alfred Young 1873–1940*, ed. G. de B. Robinson (University of Toronto Press, Toronto).

Author index